(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 April 2002 (04.04.2002)

PCT

(10) International Publication Number
WO 02/27016 A2

(51) International Patent Classification[7]: C12Q 1/00

(21) International Application Number: PCT/US01/29859

(22) International Filing Date:
25 September 2001 (25.09.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/235,268 25 September 2000 (25.09.2000) US

(71) Applicant *(for all designated States except US)*: **THE UNITED STATES OF AMERICA, REPRESENTED BY SECRETARY, DEPARTMENT OF HEALTH AND HUMAN SERVICES** [US/US]; National Institutes of Health, Office of Technology Transfer, 6011 Executive Boulevard, Suite 325, Rockville, MD 20852-3804 (US).

(71) Applicants and
(72) Inventors: **PATEL, Shailendra, B.** [GB/US]; 187 Oakpoint Landing Drive, Mount Pleasant, SC 29464 (US). **DEAN, Michael** [US/US]; 1362 Hitchingpost Lane, Frederick, MD 21703 (US).

(74) Agent: **JAY, Jeremy**; Leydig, Voit & Mayer, Ltd., Washington, 700 Thirteenth Street, Suite 300, Washington, DC 20005-3960 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/27016 A2

(54) Title: GENE INVOLVED IN DIETARY STEROL ABSORPTION AND EXCRETION AND USES THEREFOR

(57) Abstract: The present invention features ABCG5 polypeptides and ABCG5 nucleic acids, and methods of using the ABCG5 polypeptides and ABCG5 nucleic acids, for example, to identify a subject having a predisposition for developing sitosterolemia, arteriosclerosis, or heart disease; for modulating sterol transport in a cell; for modulating sterol absorption or excretion in a subject; and for identifying compounds to treat sitosterolemia.

GENE INVOLVED IN DIETARY STEROL ABSORPTION AND EXCRETION AND USES THEREFOR

CROSS-REFERENCE TO RELATED PATENT APPLICATIONS

This patent application claims the benefit of U.S. Provisional Patent Application No. 60/235,268, filed September 25, 2000, which is incorporated herein by reference.

FIELD OF THE INVENTION

This invention relates generally to identification of *ABCG5* genes that encode polypeptides involved in regulating dietary sterol absorption and excretion, and methods for using *ABCG5* nucleic acids and polypeptides.

BACKGROUND OF THE INVENTION

The molecular mechanisms that regulate the body's absorption, retention, and selective exclusion of dietary sterols, such as cholesterol and plant sterols (phytosterols), remain poorly understood. Normally, less than 5% of dietary non-cholesterol sterols are absorbed and almost none are retained. By contrast, patients suffering from sitosterolemia, a rare autosomal recessive disorder, hyper absorb and retain all sterols, including phytosterols, shellfish sterols, and cholesterol. In addition to displaying an increase in sterol absorption and loss of sterol selectivity in the intestine, patients with sitosterolemia also display impaired excretion of sterols by the liver into the bile. Consequently, these patients have highly elevated plasma phytosterol levels (in particular, sitosterol, the major plant sterol species) and develop tendon and tuberous xanthomas within the first ten years of life, as well as arthritis, accelerated arteriosclerosis, and premature coronary artery disease. Segregation analyses of these patients have shown an autosomal recessive pattern of inheritance and the sitosterolemia locus (STSL) has been mapped to chromosome 2p21, to within a 0.5 CM region. However, the precise gene defect and physiological mechanism underlying sitosterolemia has remained unknown.

Disease severity in sitosterolemia patients can sometimes be controlled by dietary restriction and administration of bile acid binding resins. Therefore, early detection of individuals with the sitosterolemia gene defect would allow earlier treatment, thereby lessening the severity of the disease. However, some individuals do not respond to current therapies. Therefore, new treatments are needed.

Epidemiological studies indicate that the incidence of breast, prostate and colon cancer are lower in communities that consume a much higher amount of plant sterols, as well as lower amounts of saturated fats. Messina and Barnes *J. Natl. Cancer. Inst.* 83:541-546 (1991). In vitro studies have established that growth of cancerous cells, such as the prostate cancer cell line LNCaP, colonic cancer cell line HT-29 and the human breast cancer cell line MDA-MB-231 can all be inhibited by exposure to sitosterol, and this can also activate apoptosis, or cell death. Mehta and Moon *Anticancer Res.* 11:593-596 (1991); Awad *et al. Anticancer Res.* 16:2797-2804 (1996); Awad *et al. Anticancer Res.* 20:821-824 (2000); Awad *et al. Int. J. Mol. Med.* 5:541-545 (2000); Awad and Fink *J. Nutr.* 130:2127-2130 (2000); Awad *et al. Nutr. Cancer* 29:212-216 (1997); Awad *et al. Nutr. Cancer* 27:210-215 (1997); and Awad *et al. Anticancer Res.* 18:471-473 (1998).

Additionally, when carcinogenic agents, such as methylnitrosourea, are fed together with high doses of sitosterol, the sitosterol supplemented animals showed reduced proliferation of cells in the intestine, with reduction of both tumors and growth retardation of tumors. Raicht *et al. Cancer Res.* 40:403-405 (1980).

Furthermore, exposure of sitosterol to cells derived from the endothelium led to an increase in the production of plasminogen activator, a beneficial agent that can lead to clearance of thrombosis. Hagiwara *et al. Thromb. Res.* 33:363-370 (1984); Shimonaka *et al. Thromb. Res.* 36:217-222 (1984). Sitosterol exposure has been shown to lead to an increased secretion of interleukin 2 and gamma interferon by activated T cells. Bouic *et al. Int. J. Immunopharmacol.* 18:693-700 (1996).

Thus, manipulating the exposure of cells to increased sitosterol levels may be beneficial for control of cancer, coronary diseases, acute thrombosis, and vascular disease. However, it is particularly beneficial that the sitosterol concentrations be kept low relative to the concentrations in sitosterolemia patients.

The present invention provides for ameliorating at least some of the deficits in the art by disclosing the gene and mutations involved in sitosterolemia, by providing the encoded polypeptides, and methods that can be used for diagnosis, treatment, and drug discovery relevant to affecting various sterol levels.

The gene involved in sitosterolemia regulates absorption of cholesterol and non-cholesterol sterols in the intestine and secretion of cholesterol and non-cholesterol sterols into the bile from the liver. Therefore, the polypeptides, nucleic acids, and methods of the invention may also be used to treat and/or prevent any disease and/or condition that would benefit from altering sterol levels systemically or locally, for example,

hypercholesterolemia, arteriosclerosis, coronary artery disease, sitosterolemia, cancers, and/or Alzheimer's disease.

BRIEF SUMMARY OF THE INVENTION

The present invention is based upon the discovery of ABCG5, a gene that encodes sterolin-1, a polypeptide involved in regulating the transport of sterols, e.g., phytosterols and shellfish sterols, and cholesterol across the cell membrane. Movement is controlled both in and out of the cell, with different affinity for sterols and cholesterol. Mutations in the *ABCG5* gene interferes with sterol transport and can cause sitosterolemia. The present invention features ABCG5 polypeptides, ABCG5 nucleic acids, and methods for regulating the activity of such polypeptides and nucleic acids, for example (but not limited to):

In accordance with an embodiment of the invention, a method of identifying a subject having a predisposition for developing sitosterolemia is provided, comprising detecting a mutant ABCG5 polypeptide or a mutated ABCG5 nucleic acid in the subject, thereby identifying a subject having a predisposition for developing sitosterolemia.

In another embodiment, a method of identifying a subject having a predisposition for developing arteriosclerosis or heart disease is provided, comprising detecting a mutant ABCG5 polypeptide or a mutated ABCG5 nucleic acid in the subject, thereby identifying a subject having a predisposition for developing arteriosclerosis or heart disease.

A method of identifying a mutant ABCG5 polypeptide or a mutated ABCG5 nucleic acid encoding the mutant polypeptide, the polypeptide having reduced selectivity for internalization of non-sterol cholesterol in an intestine or hepatic cell according to an embodiment of the invention comprises detecting, in a patient with sitosterolemia, a ABCG5 polypeptide that is not present in normal subjects or an ABCG5 nucleic acid that is not present in normal subjects, thereby identifying a mutant ABCG5 polypeptide or a mutated ABCG5 nucleic acid encoding the mutated polypeptide having reduced selectivity for internalization of non-sterol cholesterol in an intestine or hepatic cell.

In accordance with another embodiment, a method of identifying a compound for treating or preventing sitosterolemia comprises: contacting a cell culture including an ABCG5 polypeptide with a compound; and measuring ABCG5 biological activity in the cell culture, whereby an increase in ABCG5 biological activity compared to ABCG5 biological activity in a control cell culture not contacted with the compound, identifies a compound which increases ABCG5 biological activity, or, whereby a decrease in ABCG5 biological

activity compared to ABCG5 biological activity in a control cell culture not contacted with the compound, identifies a compound which decreases ABCG5 activity.

In another embodiment, a method of identifying a compound which alters ABCG5 biological activity level comprises:

contacting a mammal having cells comprising an ABCG5 polypeptide with a compound; and

measuring ABCG5 biological activity in the mammal,

whereby an increase in ABCG5 biological activity compared to ABCG5 biological activity before contacting the mammal with the compound, identifies a compound which increases ABCG5 activity, or,

whereby a decrease in ABCG5 biological activity compared to ABCG5 biological activity before contacting the mammal with the compound, identifies a compound which decreases ABCG5 activity.

An embodiment of a method of modulating transport of a sterol by a cell comprises modulating ABCG5 biological activity in the cell, thereby modulating transport of the sterol by the cell.

In another embodiment, a method of increasing sterol excretion in a subject comprises increasing ABCG5 biological activity in a hepatocyte in the subject, thereby increasing sterol excretion in the subject.

A method of decreasing sterol absorption in a subject is provided in accordance with another embodiment of the invention, comprising increasing ABCG5 biological activity in an intestinal cell in the subject, thereby decreasing sterol absorption in the subject.

In accordance with yet another embodiment, a method for improving the prognosis or ameliorating a disease state selected from the group including essentially of breast cancer, coronary heart disease, acute thrombosis, and stroke comprises administering to a patient an agent which decreases ABCG5 biological activity and results in increased sitosterol levels in said patient.

Other embodiments provided in accordance with the invention include an isolated nucleic acid encoding ABCG5, and a vector including a nucleic acid encoding ABCG5.

In accordance with other embodiments, a non-human transgenic mammal including an isolated nucleic acid encoding mammalian ABCG5, and a non-human mammal including a deleted, mutated, or polymorphic variant heterozygous ABCG5 gene, are provided.

Other embodiments provided by the invention include an isolated mammalian ABCG5 polypeptide, an isolated antibody that specifically binds an ABCG5 polypeptide, and an isolated dimer half-transporter enzyme including at least one ABCG5 monomer.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a diagram showing an amino acid sequence alignment of ABCG5 with other members of the White ABC transporter subfamily.

Fig. 2 is a diagram showing a phylogenetic tree of ABCG5-related polypeptides.

Fig. 3 is a Northern blot showing expression of the *ABCG5* gene in human tissues.

Fig. 4 is a diagram showing the pedigrees of sitosterolemia families analyzed to identify the sitosterolemia gene defect.

Fig. 5A is a diagram showing the nucleotide changes in the *ABCG5* gene in sitosterolemia patients and the resulting amino acid changes or premature polypeptide terminations.

Fig. 5B is a diagram showing a series of restriction endonuclease assays to confirm segregation of sitosterolemia mutations among family members.

Fig. 6 is a diagram showing the positions of the amino acid changes found in mutant and polymorphic variants of ABCG5.

Fig. 7 is an alignment of the human, mouse, and rat ABCG5 amino acid sequences.

Fig. 8 is a phylogenetic comparison of ABCG5 with other ABC transporter polypeptides.

Fig. 9 shows a Northern Blot of mouse mRNA from different tissues probed with a mouse ABCG5 cDNA probe.

DETAILED DESCRIPTION OF THE INVENTION

Patients with the autosomal recessive disorder sitosterolemia display elevated plasma sterol levels (particularly non-cholesterol dietary sterol) and develop tendon and tuberous xanthomas, arthritis, accelerated arteriosclerosis, and premature coronary artery disease. The present invention is based upon the identification of *ABCG5*, a novel member of the ATP-binding cassette (ABC) transporter gene family, which maps to the sitosterolemia (STSL) critical region. ABC transporter proteins bind and hydrolyze ATP to provide energy for the transport of substrates across the cell membrane. These proteins, which are divided into seven subfamilies (ABCA through ABCG), are either full size or half size; i.e., each contains either twelve transmembrane domains and two ATP-binding sites, or six

transmembrane domains and one ATP binding site. The half size molecules are believed to heterodimerize or homodimerize to form a functional transporter.

The human *ABCG5* protein contains six transmembrane domains and one ATP binding site. It contains thirteen exons and encodes a 651 amino acid, 70 kD protein, having ABC proteins characteristic motifs towards the amino-terminal end. The predicted protein is closely related to the *Drosophila white* gene and a human gene, *ABCG1*, which is induced by cholesterol. These ABC proteins all have a single ATP-binding domain at the N-terminus and a single C-terminal set of transmembrane segments. *ABCG5* maps to human chromosome 2p21, between the markers *D2S2294* and Afm210xe9. *E. J. Hum Gen* (:375-384 (2001). The expression of this gene in the liver and the intestine only suggests that the protein product has an important role in transport of specific molecule(s) into and/or out of these tissues. Its relation to sitosterolemia indicates a role in sterol absorption and non-cholesterol sterol retention, as well as impaired excretion of sterol into bile. Two different transcript sizes are detected, apparently due to alternative splicing.

While not wishing to be bound by theory, ABCG5 could either homodimerize, or heterodimerize, or exist in a state of a mixture of homodimers and heterodimers with the other known ABCG subfamily members. A possible candidate for heterodimerization partner is ABCG1, which is involved in cholesterol and phospholipid transport across the cell membrane. Klucken *et al. Proc. Nat. Acad. Sci. USA* 97:817-822 (2000).

ABCG5 may also function as a homodimer, since another ABCG subfamily member, ABCG2 (ABCP), can confer drug resistance phenotype to cells upon transfection, suggesting that it functions as a homodimer. Rabindran *et al. Cancer Res.* 60:47-50 (2000). Extra copies of ABCG5 or its partner in heterodimerization could alter the ration of homodimers/heterodimers with implication as to levels of absorption/secretion of sterols.

The *ABCG5* gene maps to the genetic interval that has been defined for sitosterolemia, for which a principal phenotype is hyper absorption of sterols by the intestine and lack of sterol transport from the liver into the bile. This leads to an accumulation of these sterols with resultant xanthomas and, in some cases, arthritis. Given that several other ABC genes play crucial roles in the transport of substances into the bile, it is likely that ABCG5 is involved in excretion of sterols from the liver. There is precedence for ABC genes playing a role in sterol transport from the finding that ABCA1 is involved in cholesterol transport from cells onto HDL. Moreover, expression of yet another member of the ABCG family, ABCG1, is induced by cholesterol loading, suggesting that ABCG1 also plays a role in cholesterol transport, presumably as a regulator of cholesterol levels

(Klucken et al., *supra*). Accordingly, stimulation of ABCG5 activity can be used to increase sterol transport from the liver into the bile, for example, to treat or prevent hypersterolemia (e.g., hypercholesterolemia or sitosterolemia) arteriosclerosis, heart disease, and/or Alzheimer's disease). Increasing ABCG5 activity can also be used to treat or prevent any other disease or condition in which it would be desirable to increase sterol transport from a cell, decrease sterol absorption by the body, or increase sterol excretion from the body.

For example, it is now known that hypercholesterolemia accelerates both beta-amyloid accumulation in the brain and Alzheimer's pathology. See, e.g., Refolo *et al. Neurobiol. Dis.* 7:321-331, (2000) and Sparks *et al. Microsc. Res. Tech.* 50:287-290, (2000). Accordingly, the methods, polypeptides, nucleic acids, and compounds of the invention can be used to decrease cholesterol absorption and/or increase cholesterol excretion to treat Alzheimer's disease and/or to prevent, ameliorate, or delay the development of Alzheimer's disease in a subject, for example, a subject at increased risk for developing the disease (e.g., a subject with hypercholesterolemia or with any other risk factor for developing Alzheimer's disease, e.g., one of the known genetic risk factors or a family history of Alzheimer's disease).

Inhibiting ABCG5 activity can be used to treat or prevent any disease or condition in which it would be desirable to decrease sterol transport by a cell, increase sterol absorption by the body, or decrease sterol excretion by the body, in a localized or systemic manner, such that an increased level of non-cholesterol sterols is observed.

For example, epidemiological studies indicate that the incidence of breast, prostate and colon cancer are lower in communities that consume a much higher amount of plant sterols, as well as lower amounts of saturated fats. Messina and Barnes, *supra.* In vitro studies have established that growth of cancerous cells, such as the prostate cancer cell line LNCaP, colonic cancer cell line HT-29 and the human breast cancer cell line MDA-MB-231 can all be inhibited by exposure to sitosterol, and this can also activate cell apoptosis. Mehta and Moon, *supra*; Awad *et al.* (1996) *supra*; Awad *et al.* (1997) *supra*; Awad *et al.* (1997) *supra*; Awad *et al.* (1998) *supra*; Awad *et al. Anticancer Res.* (2000) *supra*; Awad *et al. Int. J. Mol. Med.* (2000) *supra*; Awad *et al. Nutr. Cancer* (2000) *supra; and* Awad and Fink, *supra.* Additionally, when carcinogenic agents, such as methylnitrosourea are fed together within high doses of sitosterol, the sitosterol supplemented animals showed reduced proliferation of the cells in the intestine, with reduction of both tumors and growth retardation of tumors. Raicht *et al., supra.* Additionally, exposure of sitosterol to cells derived from the endothelium led to an increase in the production of plasminogen activator,

a beneficial agent that can lead to clearance of thrombosis. Hagiwara *et al., supra*; Shimonaka *et al., supra*. Sitosterol exposure has been shown to lead to an increased secretion of interleukin 2 and gamma interferon by activated T cells (Bouic *et al.* 1996).

Thus manipulating the exposure of cells to sitosterol can be beneficial in particular patients. Selective inhibition of ABCG5 activity leads to limited but significantly increased body level of sitosterol, which is beneficial as a chemopreventive measure for cancer, as well as for chronic inflammatory disease. Additionally, the stimulation of plasminogen activator by endothelial cells exposed to sitosterol is beneficial in acute thrombosis, such as coronary heart disease and stroke and vascular disease. A beneficial effect in these respects is the prevention, improved prognosis, or amelioration of the disease condition which is achieved when sitosterol levels are increased relative to expected pretreatment levels for that patient by at least about 5%, 10%, 20%, 30%, 50%, 70%, or 100%, preferably between about 30% to 50%.

To achieve the desired modulation of sterolin-1, one can identify and administer agents that may inhibit ABCG5 activity and lead to increased plasma and body levels of sitosterol, or conversely, agents which can increase sterolin-1 activity.

For example, possibly in combination with oral supplementation with purified phytosterols or their metabolites, or a diet rich in particular sterols, it is possible to elevate plasma levels of a desired sterol in a controlled manner. Without limiting such use or application, one example of such therapy would be to reduce the rate of growth of metastatic cancer, particularly prostate or breast cancer, and thus improve survival times for patients with these diseases. Another example would be to increase plasma sitosterol (or phytosterols and their metabolite) levels, again using agents that inhibit ABCG5 in patients with coronary heart disease and acute coronary syndromes, in whom an increase in the endothelial production of protective agents, such as plasminogen may be beneficial.

Embodiments of the present invention also provide missense and nonsense mutations in the *ABC5* gene which, when homozygous, can result in sitosterolemia. Identification of the gene and mutations involved in sitosterolemia allows genetic screening for potential carriers of the disease, as well as early identification of individuals with an increased risk for developing the disease. This ability for early detection allows earlier treatment. Additional mutations in *ABCG5* that are involved in sitosterolemia, as opposed to neutral polymorphic variations, may now be readily identified using embodiments of the invention. In addition, compounds for treating sitosterolemia or otherwise modulating sterol transport by a cell, and methods of identifying additional such compounds for treating the disease are provided by

the present invention. Furthermore, mutations in ABCG5 that cause altered sterol transport, absorption, and/or excretion, which can increase a subject's propensity for developing, e.g., hypercholesterolemia, arteriosclerosis, heart disease, and/or Alzheimer's disease, can be identified using methods described herein and/or those known in the art.

In this specification and in the claims that follow, reference is made to a number of terms which shall be defined to have the following meanings:

As used in the specification and the appended claims, the singular forms "a," "an" and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, "a molecule" can mean a single molecule or more than one molecule.

By "about" is meant ± 10% of a recited value.

By "ABCG5 biological activity" is meant any physiological function attributable to an ABCG5 polypeptide molecule, human or otherwise, including regulation of sterol (e.g., cholesterol or sitosterol) transport, absorption, or excretion by an intestinal cell and/or hepatocyte, or by any other cell expressing ABCG5 (for example, a cell transfected with a nucleic acid encoding ABCG5). ABCG5 biological activity, as referred to herein, is relative to that of the normal ABCG5 polypeptide molecule; i.e., a mutant ABCG5 polypeptide molecule, such as that produced within the body of a sitosterolemia patient, has lower than normal ABCG5 biological activity, relative to the wild type molecule. Accordingly, it will be apparent to one of ordinary skill in the art that a compound that is useful for regulating sterol transport in a cell, either *in vitro* or within a subject (e.g., in a patient in need of treatment or prevention of a disease or condition of sterol transport, such as sitosterolemia, arteriosclerosis, hypercholesterolemia, or Alzheimer's disease) will increase ABCG5 biological activity by any mechanism. However, in some cases, it may be preferable to decrease ABCG5 biological activity, as will be apparent to one of ordinary skill in the art.

Mechanisms by which a compound may increase ABCG5 biological activity include, but are not limited to, mimicry of endogenous ABCG5 polypeptide activity-mediated sterol absorption and/or excretion; stimulation of the activity of a less active or inactive version (e.g., a mutant) of the ACG5 polypeptide; or increasing the amount of ABCG5 polypeptide in a cell (e.g., by stimulating ABCG5 transcription and/or translation or by inhibiting ABCG5 mRNA or polypeptide degradation).

ABCG5 biological activity in a sample, such as a cell, tissue, or animal, may be measured using any technique for measuring sterol absorption and/or excretion by a cell, tissue, or animal, such as those described herein or known in the art. In addition, ABCG5 biological activity in a sample may be indirectly measured by measuring the relative amount

of ABCG5 mRNA (e.g., by reverse-transcription-polymerase chain reaction (RT-PCR) amplification or Northern hybridization); the level of ABCG5 polypeptide (e.g., by ELISA or Western blotting); or the activity of a reporter gene under the transcriptional regulation of an ABCG5 transcriptional regulatory region (by reporter gene assay, e.g., employing beta-galactosidase, chloramphenicol acetyltransferase (CAT), luciferase, or green fluorescent protein, as is well known in the art). For example, a compound that increases the amount of wild type ABCG5 polypeptide (or any other version of the polypeptide that maintains at least some sterol transport activity) in a cell is a compound that increases biological activity of ABCG5. In another example, a compound that increases the rate of sterol transport by a wild type, mutant, or polymorphic ABCG5 polypeptide is a compound that increases ABCG5 biological activity.

By "ABCG5 polypeptide" is meant a polypeptide that encodes an ABC half-transporter of the ABCG family that, under normal circumstances, is involved in regulating sterol absorption and/or excretion in hepatocytes. An inactivating mutation in a gene encoding an ABCG5 polypeptide can result in sitosterolemia in a subject carrying such a mutated gene. An ABCG5 polypeptide contains an amino acid sequence that bears at least 80% sequence identity, preferably at least 85% sequence identity, more preferably at least 90% sequence identity, and most preferably at least 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to a human or mouse ABCG5 polypeptide described herein.

By "wild type ABCG5 polypeptide" is meant an ABCG5 polypeptide that has normal biological activity, e.g., is produced by a normal subject not suffering from sitosterolemia. The amino acid sequence of a wild type ABCG5 polypeptide is shown in Fig. 1.

By "wild type ABCG5 nucleic acid" is meant a nucleic acid that encodes a wild type ABCG5 polypeptide.

By "polymorphic variant of an ABCG5 polypeptide" is meant an ABCG5 polypeptide containing an amino acid change, relative to wild type, that does not cause sitosterolemia. Such polymorphic amino acid variations in ABCG5 are seen in both sitosterolemia patients and in normal individuals. However, a polymorphic variant, while not the underlying cause of sitosterolemia, may subtly increase or decrease ABCG5 biological activity such that sterol transport is either more efficient or less efficient than that performed by a wild type ABCG5 polypeptide molecule.

By "mutant ABCG5 polypeptide" is meant an ABCG5 polypeptide that prematurely terminates (i.e., is not full length) or that contains an amino acid substitution such that the

polypeptide displays less biological activity than the wild type ABCG5 polypeptide, e.g., because it is less stable than the wild type polypeptide (and is thus degraded more rapidly), or because it transports less sterol than a wild type polypeptide molecule. Examples of mutant ABCG5 polypeptides are those encoded by the genes of patients suffering from sitosterolemia, as described herein.

By "mutated ABCG5 nucleic acid" is meant a nucleic acid that encodes a mutant ABCG5 polypeptide.

By "functional ABCG5 polypeptide" is meant a wild type or polymorphic ABCG5 polypeptide, or a fragment thereof, that displays sufficient biological activity to treat or prevent sitosterolemia in a subject expressing such a polypeptide.

By "test compound" is meant a molecule, be it naturally occurring or artificially derived, that is surveyed for its ability to modulate ABCG5-dependent sterol transport, absorption and/or excretion, by employing one of the assay methods described herein and/or known in the art. Test compounds may include, for example, peptides, polypeptides, synthesized organic molecules, naturally occurring organic molecules, nucleic acid molecules, and components thereof.

By "sample" is meant an animal; a tissue or organ from an animal; a cell (either within a subject, taken directly from a subject, or a cell maintained in culture or from a cultured cell line); a cell lysate (or lysate fraction) or cell extract; or a solution containing one or more molecules derived from a cell or cellular material (e.g. a polypeptide or nucleic acid), which is assayed as described herein. A sample may also be any body fluid or excretion (e.g., but not limited to, blood, urine, stool, saliva, tears, bile) that contains cells or cell components.

By "modulate" is meant to alter, by increase or decrease.

By "normal subject" is meant an individual who does not have a predisposition for developing sitosterolemia or any disease or condition involving a mutated ABCG5 gene. Such a subject typically will display a plasma phytosterol concentration of less than 1 mg/L. Salen *et al. J. Lipid Res.* 33:945-955 (1992).

By "carrier" is meant a subject who has one mutated sitosterolemia gene, but does not have a predisposition for developing the disease.

By "having a predisposition" is meant a subject who has a greater than normal chance of developing a disease or condition, such as sitosterolemia, arteriosclerosis, or heart disease, compared to the general population. Such subjects include, for example, a subject

that harbors a mutation in an ABCG5 gene such that biological activity of ABCG5 is decreased.

By an "effective amount" of a compound as provided herein is meant a nontoxic but sufficient amount of the compound to provide the desired effect, e.g., modulation of ABCG5 biological activity, for example, a decrease in sterol absorption or an increase in sterol excretion. The exact amount required will vary from subject to subject, depending on the species, age, and general condition of the subject, the severity and type of disease (or underlying genetic defect) that is being treated, the particular compound used, its mode of administration, and the like. Thus, it is not possible to specify an exact "effective amount." However, an appropriate "effective amount" may be determined by one of ordinary skill in the art using only routine experimentation.

By "pharmaceutically acceptable" is meant a material that is not biologically or otherwise undesirable, i.e., the material may be administered to an individual along with a molecule or compound of the invention (e.g., an compound that modulates ABCG5 biological activity) without causing any undesirable biological effects or interacting in a deleterious manner with any of the other components of the pharmaceutical composition in which it is contained.

By "isolated polypeptide" or "purified polypeptide" is meant a polypeptide (or a fragment thereof) that is substantially free from the materials with which the polypeptide is normally associated in nature. The polypeptides of the invention, or fragments thereof, can be obtained, for example, by extraction from a natural source (e.g., a mammalian cell), by expression of a recombinant nucleic acid encoding the polypeptide (e.g., in a cell or in a cell-free translation system), or by chemically synthesizing the polypeptide. In addition, polypeptide fragments may be obtained by any of these methods, or by cleaving full length polypeptides.

By "isolated nucleic acid" or "purified nucleic acid" is meant DNA that is free of the genes that, in the naturally-occurring genome of the organism from which the DNA of the invention is derived, flank the gene. The term therefore includes, for example, a recombinant DNA which is incorporated into a vector, such as an autonomously replicating plasmid or virus; or incorporated into the genomic DNA of a prokaryote or eukaryote (e.g., a transgene); or which exists as a separate molecule (e.g., a cDNA or a genomic or cDNA fragment produced by PCR, restriction endonuclease digestion, or chemical or *in vitro* synthesis). It also includes a recombinant DNA which is part of a hybrid gene encoding additional polypeptide sequence. The term "isolated nucleic acid" also refers to RNA, e.g.,

an mRNA molecule that is encoded by an isolated DNA molecule, or that is chemically synthesized, or that is separated or substantially free from at least some cellular components, e.g., other types of RNA molecules or polypeptide molecules.

By a "transgene" is meant a nucleic acid sequence that is inserted by artifice into a cell and becomes a part of the genome of that cell and its progeny. Such a transgene may be (but is not necessarily) partly or entirely heterologous (e.g., derived from a different species) to the cell.

By "transgenic animal" an animal comprising a transgene as described above. Transgenic animals are made by techniques that are well known in the art.

By "knockout mutation" is meant an alteration in the nucleic acid sequence that reduces the biological activity of the polypeptide normally encoded there from by at least 80% relative to the unmutated gene. The mutation may, without limitation, be an insertion, deletion, frame shift, or missense mutation. A "knockout animal," e.g., a knockout mouse, is an animal containing a knockout mutation. The knockout animal may be heterozygous or homozygous for the knockout mutation. Such knockout animals are generated by techniques that are well known in the art.

By "treat" is meant to administer a compound or molecule of the invention to a subject, such as a human or other mammal (e.g., an animal model), that has a predisposition for developing a disease or condition mediated by (or otherwise involving) high sterol levels, e.g., sitosterolemia, hypercholesterolemia, arteriosclerosis, heart disease, or Alzheimer's disease, or that has one of these diseases or conditions, in order to prevent or delay a worsening of the effects of the disease or condition (e.g., xanthomas, arthritis, arteriosclerosis, or heart disease), or to partially or fully reverse the effects of the disease. Treatment with a compound or molecule of the invention preferably increases ABCG5 biological activity sufficiently such that sterol absorption and/or excretion is altered sufficiently to halt disease progression or to allow disease reversal.

By "prevent" is meant to minimize the chance that a subject who has a predisposition for developing a disease or condition involving altered sterol transport and/or absorption and/or excretion (e.g., sitosterolemia, xanthomas, arthritis, hypercholesterolemia, arteriosclerosis, heart disease, or Alzheimer's disease) will develop the disease or condition. For example, a compound that prevents the development of sitoserolemia will increase ABCG5 biological activity in the subject such that manifestations of the disease are minimized or avoided.

By "specifically binds" is meant that an antibody recognizes and physically interacts with its cognate antigen (i.e., an ABCG5 polypeptide) and does not significantly recognize and interact with other antigens; such an antibody may be a polyclonal antibody or a monoclonal antibody, which are generated by techniques that are well known in the art.

By "probe," "primer," or oligonucleotide is meant a single-stranded DNA or RNA molecule of defined sequence that can base-pair to a second DNA or RNA molecule that contains a complementary sequence (the "target"). The stability of the resulting hybrid depends upon the extent of the base-pairing that occurs. The extent of base-pairing is affected by parameters such as the degree of complementarity between the probe and target molecules and the degree of stringency of the hybridization conditions. The degree of hybridization stringency is affected by parameters such as temperature, salt concentration, and the concentration of organic molecules such as formamide, and is determined by methods known to one skilled in the art. Probes or primers specific for ABCG5 nucleic acids (e.g., genes and/or mRNAs) have at least 80%-90% sequence complementarity, preferably at least 91%-95% sequence complementarity, more preferably at least 96%-99% sequence complementarity, and most preferably 100% sequence complementarity to the region of the ABCG5 nucleic acid to which they hybridize. Probes, primers, and oligonucleotides may be detectably-labeled, either radioactively, or non-radioactively, by methods well-known to those skilled in the art. Probes, primers, and oligonucleotides are used for methods involving nucleic acid hybridization, such as: nucleic acid sequencing, reverse transcription and/or nucleic acid amplification by the polymerase chain reaction, single stranded conformational polymorphism (SSCP) analysis, restriction fragment polymorphism (RFLP) analysis, Southern hybridization, Northern hybridization, *in situ* hybridization, electrophoretic mobility shift assay (EMSA).

By "specifically hybridizes" is meant that a probe, primer, or oligonucleotide recognizes and physically interacts (i.e., base-pairs) with a substantially complementary nucleic acid (e.g., an ABCG5 nucleic acid of the invention) under high stringency conditions, and does not substantially base pair with other nucleic acids.

By "high stringency conditions" is meant conditions that allow hybridization comparable with that resulting from the use of a DNA probe of at least 40 nucleotides in length, in a buffer containing 0.5 M $NaHPO_4$, pH 7.2, 7% SDS, 1 mM EDTA, and 1% BSA (Fraction V), at a temperature of 65°C, or a buffer containing 48% formamide, 4.8X SSC, 0.2 M Tris-Cl, pH 7.6, 1X Denhardt's solution, 10% dextran sulfate, and 0.1% SDS, at a temperature of 42°C. Other conditions for high stringency hybridization, such as for PCR,

Northern, Southern, or *in situ* hybridization, DNA sequencing, etc., are well-known by those skilled in the art of molecular biology. See, e.g., F. Ausubel *et al. Current Protocols in Molecular Biology*, John Wiley & Sons, New York, NY, 1998, hereby incorporated by reference.

Identification of compounds that affect ABCG5-mediated sterol absorption or excretion

ABCG5 is normally highly expressed in the liver and in the intestine. Sitosterolemia patients with mutations in the ABCG5 gene show hyper absorption of sterols (e.g., cholesterol and sitosterol) in the intestine and decreased sterol excretion into bile acids within the liver. Accordingly, wild type, polymorphic, and mutant ABCG5 polypeptides, and the nucleic acids encoding these polypeptides, may be employed in various types of high-throughput screening assays for identification of compounds that inhibit sterol absorption and/or stimulate sterol excretion in an ABCG5-dependent manner. Such compounds are useful for treating and/or preventing sitosterolemia, hypercholesterolemia, arteriosclerosis, heart disease, and any other disease of sterol accumulation (e.g., the toxic excess of cholesterol in the brains of Alzheimer's patients).

A) Lipid transport assays

For example, to identify a compound that inhibits sterol absorption and/or stimulates sterol excretion from cells, a nucleic acid encoding a wild type, polymorphic, or mutant ABCG5 can be stably or transiently transfected into an established cultured cell line that does not normally express ABCG5 (e.g., human 293 cells or Chinese Hamster Ovary (CHO) cells). In one example, to isolate cells that stably express ABCG5, a nucleic acid encoding ABCG5 is inserted into an expression plasmid, under the transcriptional regulation of a eukaryotic promoter such as the CMV or RSV promoter. Cells containing the plasmid are identified and/or selected by well-known techniques, for example, by using an expression plasmid that also allows for co-expression of a selectable marker, such as an antibiotic resistance gene. Drug-resistant cells can be cloned and ABCG5 expression can be confirmed, e.g., by RT-PCR, Northern blotting, ELISA, or Western blotting. Once an appropriate cell clone has been identified, it can be used in sterol absorption/excretion assays to identify compounds that regulate this process in an ABCG5-dependent fashion. Such a cell line can be conveniently grown in a multi-well format and exposed to a library of compounds in the presence of labeled cholesterol, sitosterol, or another sterol.

In a general example, cells expressing wild type, polymorphic, or mutant ABCG5 are cultured in the presence of a labeled sterol, e.g., radiolabeled cholesterol or radiolabeled sitosterol, or a sterol fluorophore, such as fluoresterol, which is used to trace cholesterol absorption. Detmers *et al. Biochim. Biophys. Acta.* 1486:243-252 (2000); Hernandez *et al., Biochim. Biophys. Acta.* 1486:232-242 (2000); and Sparrow *et al. J. Lipid Res.* 40:1747-1757 (1999). The cells are incubated with the labeled sterol in the presence and absence of a test compound, after which the intracellular concentrations of sterol in the presence versus the absence of the test compound are compared. A test compound that decreases intracellular sterol concentrations, relative to intracellular sterol concentrations in control cells not treated with the test compound, is a compound that decreases sterol absorption and/or increases sterol excretion. One of ordinary skill in the art will understand that compounds that preferentially affect the absorption and/or excretion of a particular sterol, e.g., cholesterol versus sitosterol, may be readily identified by performing parallel measurements, in separate cell samples, of the relative effect of the test compound on the absorption/excretion of cholesterol versus sitosterol. A compound that preferentially regulates sterol absorption/excretion, for example, may be useful for treating and/or preventing hypersterolemia (e.g., sitosterolemia or hypercholesterolemia), arteriosclerosis, heart disease, and/or Alzheimer's disease in patients that are prone to such conditions, e.g., due to an ABCG5 gene defect or another type of genetic or physiological defect (e.g., morbid obesity).

Such screening assays can also be performed using cell lines that naturally express ABCG5 and provide a model for intestinal absorption/excretion of sterols, for example, human CaCo2 cells grown under polarized conditions. Field *et al. J. Lipid Res.* 24:409-417 (1983); Field *et al. J. Lipid Res.* 38:348-360 (1997). As described above, cells are incubated with labeled sterol, in the presence and absence of the test compound. The ability of the compound to inhibit sterol uptake or stimulate sterol excretion by the cells allows the identification of compounds that can be further tested for specificity and potency by techniques that are known to one of ordinary skill in the art. For example, a compound intended to control plasma cholesterol or sitosterol levels by either inhibiting cholesterol or sitosterol absorption in the gut or stimulating cholesterol or sitosterol excretion by the liver may be tested in laboratory animals, such as mice, that contain normal, polymorphic, mutated, or deleted ABCG5 genes. Plasma levels of the sterol of interest are measured in treated or untreated animals.

As will be recognized by one of ordinary skill in the art, there are numerous modifications that can be made to the basic assay. For example, intestinal cells or hepatocytes, which normally express ABCG5, may be used in the assays of the invention. These cells may be obtained from normal individuals or from individuals with sitosterolemia.

In another variation, if hepatocytes are being used in a screening assay, donor molecules, such as high density lipoproteins (HDL), which are known to promote efficient flux of cholesterol between plasma and hepatocytes. Robins *et al. Hepatology* 29:1541-1548 (1999); Robins *et al. J Clin. Invest.* 99:380-384 (1997) may be added to the cells along with the labeled sterol and test compound. Under these circumstances, the transfer of the sterol into the cell from the HDL is matched by efflux, governed by the activity of ABCG5. Any test compound that can attenuate or stimulate this process may be useful for therapeutic modulation of sterol absorption and/or excretion.

Competition assays, for example, using photo-activatable sterols, can also be used to identify compounds that modulate (increase or decrease) binding of cholesterol, sitosterol, and/or other sterols, to ABCG5, and thus can be used to modulate sterol absorption and excretion by intestinal cells and/or liver cells.

A protein fragment of another ABC transporter protein is, for example, a type of compound that can be an agent for modulation of ABCG5 activity, most likely for reducing activity. Because ABCG5 is now understood to form a heterodimer with another "half-transporter," i.e., a six trans-membrane domains transporter, fragments of such transporter protein can compete with the ABCG5 partner for dimerization with ABCG5. For example, fragments of ABCI compete with ABC1 for formation of heterodimer with ABCG5. The fragments for testing can be introduced into a cell culture as peptides, or could be expressed within a test cell engineered to express particular such fragments.

An antibody or active antibody fragment which specifically binds sterolin-1 protein can be an antagonist of ABCG5 activity. Similarly, an antibody or active antibody fragment which binds a proposed heterodimer partner to ABCG5, for example, another half-transporter ABC transporter protein, or an Ab_2 (anti-Id Ab) specific to an antibody or active antibody fragment which binds a proposed heterodimer partner to ABCG5, can be an antagonist of sterolin-1 activity. Furthermore, particular antibodies can be specific to mutated ABCG5 polypeptides and help recognize them. This can be very useful when the Ab Id used as part of a prognosis or diagnosis in which the presence of mutated ABCG5 polypeptide is detected. Methods to raise antibodies are well known in the art. Initially,

polyclonal antibodies can be raised and tested in, for example, an in-vitro assay. Such an assay can involve, for example, an assessment of sterol movement into or out of cells in the presence of a sera shown *in vitro* to be an antibody specific to ABCG5 or ABCG5 dimerization partner, as discussed above. Eventually, candidate Abs could be developed into monoclonal antibodies.

B) Transcription regulation of ABCG5 expression

Another method for reducing or preventing elevated plasma cholesterol or sitosterol levels (a risk factor for heart, stroke, and atherosclerotic disease) is to decrease sterol absorption in the intestine and/or increase sterol excretion by the liver by increasing ABCG5 expression. The promoter of the ABCG5 gene can be used to identify factors that regulate, i.e., increase or decrease, ABCG5 gene transcription. Precedent for such therapeutic transcriptional regulation is found in the identification of drugs such as the thiozolidinedione compounds used to treat diabetes and fibric acid derivatives to treat lipid disorders. Similarly, the ABCG5 transcriptional promoter can be used to identify important transcription factors and DNA motifs that can be targeted to up-regulate ABCG5 gene transcription, leading to increased ABCG5 biological activity. Further yet, the known sequence of ABCG5 mRNA can allow for design of mRNA destabilizers, such as antisense constructs, ribozymes, or co-transcriptional repressor constructs, as known in the art.

Screening assays for compounds that transcriptionally regulate the ABCG5 gene are performed using cells or animals containing an episomal or stably integrated chimeric plasmid construct that contains the ABCG5 promoter region driving expression of a nucleic acid encoding ABCG5 or a reporter gene product such as green fluorescent protein, alkaline phosphatase, chloramphenicol acetyltransferase, luciferase, and beta-galactosidase. Expression of ABCG5 or the reporter gene product by a cell expressing such a construct is compared in the presence and absence of the test compound. Compounds that increase or decrease ABCG5 promoter activity can then be readily identified and further characterized.

Transcription factors that regulate activity of the ABCG5 gene can be identified using well known techniques, for example, but not limited to, gel shift assays, DNAse protection assays, and reporter gene assays. Any transcription factor so identified can itself be used as a potential therapeutic target in assays to identify therapeutic compounds for modulating ABCG5 biological activity. Compounds that directly or indirectly modulate transcription of the ABCG5 gene are useful for regulating sterol transport, absorption, and/or excretion at the cellular level and/or whole-body level, and therefore, are useful for

treating, ameliorating, and/or preventing any disease or condition in which it would be beneficial to modulate transport, absorption, and/or excretion of sterols or to otherwise regulate lipid levels (e.g., high density lipoprotein cholesterol (HDL-C), low density lipoprotein cholesterol (LDL-C), and/or triglycerides); such diseases and conditions include, e.g., sitosterolemia, arteriosclerosis, and cardiovascular disease.

For example, a compound that inhibits the activity of a transcriptional repressor of the ABCG5 gene would up-regulate expression of ABCG5 and therefore increase ABCG5 biological activity; such a compound can be used to inhibit sterol absorption by the intestine and/or increase sterol excretion by the liver. A compound that stimulates activity of an ABCG5 transcriptional activator would also increase ABCG5 expression, and therefore, also can be used to inhibit sterol absorption by the intestine and/or increase sterol excretion by the liver. In yet another example, stimulation of ABCG5 expression in atheroscleotic plaques (for example, by stimulating ABCG5 expression in macrophages) could be used to effect sterol efflux from such plaques, thereby resulting in plaque stabilization and regression.

Alternatively, transcriptional factors which reduce the activity of ABCG5 can be useful agents for increasing sterol levels in a patient. Such factors are agents which bind to the DNA region upstream of the ABCG5 gene.

Transcription factors known to regulate apolipoprotein genes or other cholesterol- or lipid-regulating genes are of particular relevance in screens for the discovery of compounds that regulate activity of the ABCG5 gene. Such transcription factors include, but are not limited to, the steroid response element binding proteins (SREBP-1 and SREBP-2), and the PPAR (peroxisomal proliferation-activated receptor), RXR, FXR (farnesoid X receptor) and LXR (liver X receptor) transcription factors (Horton *et al.* Curr. Opin. Lipidol. 10:143-150 (1999); Brown *et al. Nutr. Rev.* 56:S1-3 (1998); Buchan *et al. Med. Res. Rev.* 20:350-366 (2000); Rosen *et al. Genes Dev.* 14:1293-1307 (2000); Gervois *et al. Clin. Chem. Lab. Med.* 38:3-11 (2000); Forman *et al. Proc. Nat. Acad. Sci. U.S.A.* 94:10588-10593 (1997); Schroepfer, *Physiol. Rev.* 80:361-554 (2000); Mangelsdorf *et al. Cell* 83:841-850 (1995). For example, LXRs may alter transcription of ABCG5 by mechanisms involving heterodimerization with retinoid X receptors (RXRs) and then binding to specific response elements (LXREs). Examples of such LXRs include LXRα and LXRβ (Mangelsdorf *et al. Cell* 83:841-850 (1995); and Repa *et al. Science* 289:1524-1529 (2000). Janowski *et al. Proc. Natl. Acad. Sci. USA* 96:266-271 (1999) describes the role of naturally occurring oxysterols in LXR-dependent transactivation through the promoter for cholesterol 7α-

hydoxylase (Cyp7a), which is the rate limiting enzyme in bile acid synthesis, and demonstrates that oxysterols bind directly to LXRs. Compounds that modulate LXR-mediated transcriptional activation are likely to modulate ABCG5 gene expression and thus are useful for modulating sterol absorption and excretion. Repa *et al. Science* 289:1524-1529 (2000).

Compounds known to modulate LXR activity include, without limitation, 24-(S),25-epoxycholesterol; 24(S)-hydroxycholesterol; 22-(R)-hydroxycholesterol; 24(R),25-epoxycholesterol; 22(R)-hydroxy-24(S),25-epoxycholesterol; 22(S)-hydroxy-24(R),25-epoxycholesterol; 24-(S),25-iminocholesterol; methyl-38-hydroxycholonate; N,N-dimethyl-3β-hydroxycholonamide; 24(R)-hydroxycholesterol; 22(S)-hydroxycholesterol; 22(R),24(S)-dihydroxycholesterol; 25-hydroxycholesterol; 22(R)-hydroxycholesterol; 22(S)-hydroxycholesterol; 24(S),25-dihydroxycholesterol; 24(R),25-dihydroxycholesterol; 24,25-dehydrocholesterol; 25-epoxy-22(R)-hydroxycholesterol; 20(S)-hydroxycholesterol; (20R,22R)-cholest-5-ene-3β,20,22-triol; 4,4-dimethyl-5-α-cholesta-8,14,24-trien-3-β-ol; 7α-hydroxy-24(S),25-epoxycholesterol; 7α-hydroxy-24(S),25-epoxycholesterol; 7-oxo-24(S),25-expoxycholesterol; 7α-hydroxycholesterol; 7-oxocholesterol; and desmosterol. Additional LXR-modulating compounds are described, for example, in Janowski *et al. Nature* 383:728-731 (1996); Lehman *et al. J. Biol. Chem.* 272:3137-3140 (1997); and Janowski *et al. Proc. Natl. Acad. Sci. USA* 96:266-271 (1998), each of which is herein incorporated by reference in its entirety. In addition, one of ordinary skill in the art will recognize that synthetic sterols having LXR-modulating activity can be readily identified using screening methods known in the art (see, for example, Janowski *et al. Proc. Natl. Acad. Sci. USA* 96:266-271 (1998). Non-steroidal agonists such as RIP140 protein, antibodies (monoclonal or polyclonal) specific for LXRα or LXRβ; tetradecycloxy-furnacarboxylic acid (TOFA;); tetradecylthioacetic acid; as well as other fatty acids (see, for example, Tobin *et al. Molec. Endocrin.* 14:741-752 (2000) are also useful LXR-modulating agents and can be used to identify compounds that are useful in the methods of the present invention.

Additional transcription factors which may also be useful for modulating ABCG5 gene expression, and thereby cellular and/or whole-body transport, absorption, and/or excretion of sterols, include REV-ERB-, SREBP-1 & 2, ADD-1, EBPα, CREB binding protein, P300, HNF 4, RAR, and RORα (Horton *et al. Curr. Opin. Lipidol.* 10:143-150 (1999); Brown *et al. Nutr. Rev.* 56:S1-3 (1998); Buchan *et al. Med. Res. Rev.* 20:350-366 (2000); Rosen *et al. Genes Dev.* 14:1293-1307 (2000); Gervois *et al. Clin. Chem. Lab. Med.*

38:3-11 (2000); Forman *et al. Proc. Nat. Acad. Sci. U.S.A.* 94:10588-10593 (1997); Schroepfer, *Physiol. Rev.* 80:361-554 (2000); Mangelsdorf *et al. Cell* 83:841-850 (1995; and Forman *et al. Molec. Endocrinol.* 8:1253-1261 (1994). RXR heterodimerizes with many nuclear receptors, including LXR, and aids in transactivating the target gene. Thus, compounds that modulate RXR-mediated transcriptional activity will also modulate ABCG5 expression. Numerous RXR-modulating compounds (rexinoid compounds, see, e.g., Liu *et al. Int. J. Obes. Relat. Metab. Disord.* 24:997-1004(2000) are known in the art, including, for example, hetero ethylene derivatives; tricyclic retinoids; trienoic retinoids; benzocycloalkenyl-alka:di- or trienoic acid derivatives; bicyclic-aromatic compounds and their derivatives; bicyclylmethyl-aryl acid derivatives; phenyl-methyl heterocyclic compounds; tetrahydro-napthyl compounds; arylthio-tetrahydro-naphthalene derivatives and heterocyclic analogues; 2,4-pentadienoic acid derivatives; tetralin-based compounds; nonatetraenoic acid derivatives; SR11237; dexamethasone; hydroxy, epoxy, and carboxy derivatives of methoprene; bicyclic benzyl, pyridinyl, thiophene, furanyl, and pyrrole derivatives; benzofuran-acrylic acid derivatives; aryl-substituted and aryl and (3-oxo-1-propenly)-substituted benzopyran, benzothiopyran, 1,2-dihydroquinoline, and 5,6-dihydronaphthalene derivatives; vitamin D3 (1,25-dihydroxyvitamin D3) and analogs; 24-hydroxylase inhibitor; mono-or polyenic carboxylic acid derivatives; tetrahdroquinolin-2-one-6 or 7-yl and related derivatives; tetrahydronaphthalene; oxyiminoalkanoic acid derivatives; LG 100268; and LGD 1069. Additional compounds include BRL 49653; troglitazone; pioglitazone; ciglitazone; WAY-120; englitazone; AD 5075; and darglitazone.

Compounds found to be effective at modulating the level of cellular ABCG5 expression may be confirmed as useful in animal models (for example, mice, rats, pigs, rabbits, or chickens; see, e.g., Smith, JD, *Lab. Anim. Sci.* 48:573-579 (1998); Narayanaswamy *et al. J. Vasc. Interv. Radiol.* 11:5-17 (2000); Poernama *et al. Ateriosclerosis. Thromb.* 12:601-607 (1992); and Schreyer *et al. Ateriosclerosis. Thromb.* 14:2053-2059 (1994). For example, a useful compound may ameliorate absorption of dietary cholesterol by the intestine, or increase excretion of cholesterol into bile. A compound that promotes an increase in ABCG5 expression or activity is considered particularly useful in the invention; such a molecule may be used, for example, as a therapeutic to increase the level or activity of native, cellular ABCG5 and thereby lower plasma cholesterol levels in an animal (for example, a human).

Animal Models

Compounds identified as modulating ABCG5 expression may be subsequently screened in any available animal model system, including, but not limited to, mice, rats, pigs, rabbits, and chickens. Smith, JD *Lab. Anim. Sci.* 48:573-579 (1998); Narayanaswamy *et al. J. Vasc. Interv. Radiol.* 11:5-17 (2000); Poernama *et al. Aterioscler. Thromb.* 12:601-607 (1992); and Schreyer *et al. Aterioscler. Thromb.* 14:2053-2059 (1994). Test compounds are administered to these animals according to standard methods.

Animal models that mimic diseases and conditions involving ABCG5-dependent alterations in transport, absorption, and or excretion are known in the art and/or can be developed using conventional molecular biology methods. For example, a transgenic animal (e.g., a mouse) that over-expresses ABCG5 in its liver or intestine can be generated by inserting an ABCG5-encoding nucleic acid under the transcriptional regulation of the appropriate tissue-specific promoter into the genome of the animal. For example, when the ABCG5 cDNA is placed under transcriptional regulation of the fatty acid binding protein promoter (Sweetser *et al. J. Biol Chem.* 262:16060-16071 (1987); Sweetser *et al. Proc. Nat. Acad. Sci. USA* 85:9611-9615 (1988), expression is confined to the intestine. In another example, placing the ABCG5 cDNA under the CD68 promoter (Greaves *et al. Genomics* 54:165-168 (1998) results in high levels of expression in macrophages. Such transgenic animals are then made hyperlipidemic (e.g., by cross-breeding them to apoE knock-out mice or, by providing them with a diet or administering a drug that stimulates hyperlipidemia) to test whether the atherosclerotic process can be ameliorated by over-expression of ABCG5, or by administering a compound that stimulates ABCG5 biological activity.

Provided herein are the mouse, and partial rat and hamster ABCG5 cDNA sequence and genomic location/organization of the mouse Abcg5 gene equivalent, as well as the mouse ABCG5 polypeptides, and rat cDNA and polypeptide sequences. Accordingly, knockout mice devoid of an active copy of the natural mouse ABCG5 can be constructed and used in assaying ABCG5 constructs and agents for modulation of their activities, by methods well known in the art. Additionally the mouse or rat ABCG5 nucleic acids and polypeptides can be an alternative source of materials and constructs for identifying modulators of human ABCG5 activity.

Test Compounds

In general, novel drugs that modulate sterol (e.g., cholesterol or phytosterols) absorption and/or excretion by modulating ABCG5 biological activity may be identified from large libraries of natural products or synthetic (or semi-synthetic) extracts or chemical

libraries according to methods known in the art. Those skilled in the field of drug discovery and development will understand that the precise source of test extracts or compounds is not critical to the screening procedure(s) of the invention. Accordingly, virtually any number of chemical extracts or compounds can be screened using the exemplary methods described herein. Examples of such extracts or compounds include, but are not limited to, plant-, fungal-, prokaryotic- or animal-based extracts, fermentation broths, and synthetic compounds, as well as modification of existing compounds. Numerous methods are also available for generating random or directed synthesis (e.g., semi-synthesis or total synthesis) of any number of chemical compounds, including, but not limited to, saccharide-, lipid-, peptide-, and nucleic acid-based compounds. Synthetic compound libraries are commercially available, e.g., from Brandon Associates (Merrimack, NH) and Aldrich Chemical (Milwaukee, WI). Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant, and animal extracts are commercially available from a number of sources, including Biotics (Sussex, UK), Xenova (Slough, UK), Harbor Branch Oceangraphics Institute (Ft. Pierce, FL), and PharmaMar, U.S.A. (Cambridge, MA). In addition, natural and synthetically produced libraries are generated, if desired, according to methods known in the art, e.g., by standard extraction and fractionation methods. Furthermore, if desired, any library or compound is readily modified using standard chemical, physical, or biochemical methods.

In addition, those skilled in the art of drug discovery and development readily understand that methods for dereplication (e.g., taxonomic dereplication, biological dereplication, and chemical dereplication, or any combination thereof) or the elimination of replicates or repeats of materials already known for their ABCG5-modulatory activities should be employed whenever possible.

When a crude extract is found to modulate ABCG5-dependent sterol absorption and/or excretion, further fractionation of the positive lead extract is necessary to isolate chemical constituents responsible for the observed effect. Thus, the goal of the extraction, fractionation, and purification process is the careful characterization and identification of a chemical entity within the crude extract having an activity that mimics, stimulates, or antagonizes ABCG5, depending upon the effect desired. The same assays described herein for the detection of activities in mixtures of compounds can be used to purify the active component and to test derivatives thereof. Methods of fractionation and purification of such heterogenous extracts are known in the art. If desired, compounds shown to be useful agents for treatment are chemically modified according to methods known in the art.

Compounds identified as being of therapeutic value can be subsequently analyzed using any standard animal models for a disease or condition in which it is desirable to regulate ABCG5-modulated sterol absorption and/or excretion (e.g., sitosterolemia, hypercholesterolemia, arteriosclerosis, heart disease, and/or Alzheimer's disease), as described herein.

Administration of compounds that modulate ABCG5 biological activity

The compositions and methods described herein can be used therapeutically in combination with a pharmaceutically acceptable carrier. By "pharmaceutically acceptable carrier" is meant a material that is not biologically or otherwise undesirable, i.e., the material may be administered to an individual along with a polypeptide, nucleic acid, or other compound of the invention without causing any undesirable biological effects or interacting in a deleterious manner with any of the components of the pharmaceutical composition in which it is contained. Pharmaceutical carriers are well-known in the art. These most typically are standard carriers for administration of vaccines or pharmaceuticals to humans, including solutions such as sterile water, saline, and buffered solutions at physiological pH.

Molecules intended for pharmaceutical delivery may be formulated in a pharmaceutical composition. Pharmaceutical compositions may include carriers, thickeners, diluents, buffers, preservatives, surface active agents and the like in addition to the molecule of choice. Pharmaceutical compositions may also include one or more active ingredients such as antimicrobial agents, anti-inflammatory agents, anesthetics, and the like. Methods for making such formulations are well known in the art, and are described, for example, in: *Remington:* THE SCIENCE AND PRACTICE OF PHARMACY (19th ed.), ed. A.R. Gennaro, E.W. Martin Mack Publishing Co., Easton, PA, 1995.

The pharmaceutical compositions may be administered in a number of ways depending on whether local or systemic treatment is desired, and on the area to be treated. Administration may be topically (including ophthalmically, vaginally, rectally, intranasally), orally, by inhalation, or parenterally, for example by intravenous drip, subcutaneous, intraperitoneal or intramuscular injection. The compounds and compositions of the present invention can be administered intravenously, intraperitoneally, intramuscularly, subcutaneously, intracavity, or transdermally.

Preparations for parenteral administration include sterile aqueous or non-aqueous solutions, suspensions, and emulsions. Examples of non-aqueous solvents are propylene

glycol, polyethylene glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyl oleate. Aqueous carriers include water, alcoholic/aqueous solutions, emulsions or suspensions, including saline and buffered media. Parenteral vehicles include sodium chloride solution, Ringer's dextrose, dextrose and sodium chloride, lactated Ringer's, or fixed oils. Intravenous vehicles include fluid and nutrient replenishers, electrolyte replenishers (such as those based on Ringer's dextrose), and the like. Preservatives and other additives may also be present such as, for example, antimicrobials, anti-oxidants, chelating agents, and inert gases and the like.

Formulations for topical administration may include ointments, lotions, creams, gels, drops, suppositories, sprays, liquids and powders. Conventional pharmaceutical carriers, aqueous, powder or oily bases, thickeners and the like may be necessary or desirable.

Compositions for oral administration include powders or granules, suspensions or solutions in water or non-aqueous media, capsules, sachets, or tablets. Thickeners, flavorings, diluents, emulsifiers, dispersing aids or binders may be desirable. Formulations for parenteral administration may include sterile aqueous solutions which may also contain buffers, diluents and other suitable additives.

The compounds of the invention are administered in an effective amount, using standard approaches. By "effective amount" is meant the amount of compound that is useful for performing its stated function, e.g., inhibiting ABCG5-mediated sterol absorption and/or stimulating ABCG5-mediated sterol excretion in the intestine and/or liver. Effective dosages and schedules for administering the compounds may be determined empirically, and making such determinations is routine to one of ordinary skill in the art. The skilled artisan will understand that the dosage will vary, depending upon, for example, the species of the subject the route of administration, the particular compound to be used, other drugs being administered, and the age, condition, sex and extent of the disease in the subject. The dosage can be adjusted by the individual physician in the event of any counterindications. A dose of a compound of the invention generally will range between about 1 μg/kg of body weight and 1 g/kg of body weight. Examples of such dosage ranges are, e.g., about 1 μg-100 μg/kg, 100 μg/kg-10 mg/kg, or 10 mg-1 g/kg, once a week, bi-weekly, daily, or two to four times daily. Compounds of the invention include ABCG5 polypeptides, ABCG5 nucleic acids, and molecules that regulate expression and/or biological activity of endogenous wild type, polymorphic, and/or mutant ABCG5 polypeptides and/or nucleic acids (e.g., DNA or RNA molecules) encoding such ABCG5 polypeptides.

Nucleic Acid Delivery

ABCG5 biological activity can be stimulated in a subject by administering to the subject a nucleic acid encoding ABCG5, using any method known for nucleic acid delivery into the cells of a subject. The ABCG5 nucleic acid is taken up by the cells of the subject and directs expression of the encoded ABCG5 in those cells that have taken up the nucleic acid. The ABCG5 nucleic acids of the present invention can be in the form of naked DNA or RNA, or the nucleic acids can be within a vector for delivering the nucleic acids to the cells. The vector can be a commercially available preparation, such as an adenovirus vector (Quantum Biotechnologies, Inc. (Laval, Quebec, Canada). Delivery of the nucleic acid or vector to cells can be via a variety of mechanisms. As one example, delivery can be via a liposome, using commercially available liposome preparations such as LIPOFECTIN, LIPOFECTAMINE (GIBCO-BRL, Inc., Gaithersburg, MD), SUPERFECT (Qiagen, Inc. Hilden, Germany) and TRANSFECTAM (Promega Biotec, Inc., Madison, WI), as well as other liposomes developed according to procedures standard in the art. In addition, the nucleic acid or vector of this invention can be delivered *in vivo* by electroporation, the technology for which is available from Genetronics, Inc. (San Diego, CA) as well as by means of a SONOPORATION machine (ImaRx Pharmaceutical Corp., Tucson, AZ).

As one example, vector delivery can be via a viral system, such as a retroviral vector system which can package a recombinant retroviral genome. See e.g., Pastan *et al. Proc. Natl. Acad. Sci. U.S.A.* 85:4486 (1988); Miller *et al. Mol. Cell. Biol.* 6:2895 (1986). The recombinant retrovirus can then be used to infect and thereby deliver to the infected cells a nucleic acid that encodes an ABCG5 polypeptide. The exact method of introducing the altered nucleic acid into mammalian cells is, of course, not limited to the use of retroviral vectors. Other techniques are widely available for this procedure including the use of adenoviral vectors (Mitani *et al. Hum. Gene Ther.* 5:941-948 (1994), adeno-associated viral (AAV) vectors (Goodman *et al. Blood* 84:1492-1500 (1994), lentiviral vectors (Naidini *et al. Science* 272:263-267 (1996), pseudotyped retroviral vectors (Agrawal *et al. Exper. Hematol.* 24:738-747 (1996). Physical transduction techniques can also be used, such as liposome delivery and receptor-mediated and other endocytosis mechanisms. See, for example, Schwartzenberger *et al. Blood* 87:472-478 (1996). The present invention can be used in conjunction with any of these or other commonly used gene transfer methods.

In a particular example, to deliver an ABCG5 nucleic acid to the cells of a human subject in an adenovirus vector, the dosage can range from about 10^7 to 10^9 plaque forming unit (pfu) per injection but can be as high as 10^{12} pfu per injection. Crystal, *Hum. Gene*

Ther. 8:985-1001 (1997); Alvarez and Curiel, *Hum. Gene Ther.* 8:597-613 (1997). Ideally, a subject will receive a single injection. If additional injections are necessary, they can be repeated at six month intervals for an indefinite period and/or until the efficacy of the treatment has been established.

Parenteral administration of the nucleic acid or vector of the present invention, if used, is generally characterized by injection. Injectables can be prepared in conventional forms, either as liquid solutions or suspensions, solid forms suitable for solution of suspension in liquid prior to injection, or as emulsions. A more recently revised approach for parenteral administration involves use of a slow release or sustained release system such that a constant dosage is maintained. See, e.g., U.S. Patent No. 3,610,795, which is incorporated by reference herein. For additional discussion of suitable formulations and various routes of administration of therapeutic compounds, see, e.g., *Remington: The Science and Practice of Pharmacy* (19th ed.) ed. A.R. Gennaro, Mack Publishing Company, Easton, PA 1995.

The present invention is more particularly described in the following examples which are intended as illustrative only since numerous modifications and variations thereof will be apparent to those of ordinary skill in the art.

Example I: A liver-specific ATP-binding cassette gene (*ABCG5*) from the ABCG (White) gene subfamily maps to human chromosome 2p21 in the region of the sitosterolemia locus

Methods

RNA Expression Analysis

Labeling of cDNAs and of individual probes was accomplished using the Rediprime II random prime labeling system according to the manufacturer's instructions (Amersham, Arlington Heights, Il). Probes were hybridized to multiple tissue Northern blots from Clontech (Palo Alto, CA) according to the manufacturers protocol. A quantitative real-time PCR analysis assay was developed for ABCG5 and several other ABC genes using the Cyber-green expression system (Perkin-Elmer, Foster City, CA).

cDNA, genomic cloning and exon/intron structure

Primers were designed from the sequence of the EST clones and used for the amplification of White3 gene fragments from a fetal liver cDNA library (Clontech). Primers White3 RACE3c (5'-AGTCGGTCTGCCACATGGCTCAGACTC) and White RACE4 (5'-CGCAGCGCCCGGCCGTTCACATACACC) were used for 5' RACE

reactions using Marathon-Ready cDNA (Clontech). PCR products were cloned into the pCR2.1-TOPO vector (Invitrogen). Primers for amplification of genomic fragments were designed from White3 cDNA sequence. Platinum Taq DNA Polymerase High Fidelity (GibcoBRL) was used for Long Range PCR. The positions of the introns were determined by comparison between genomic and cDNA sequences. Primers for amplification of individual exons were designed from adjacent intron sequence 30-50 base pairs (bp) from the splice site. Amplification of exons was performed with AmpliTaq Gold Polymerase (Perkin Elmer) according to protocol. Sequencing was performed with DNA Sequencing Kit (Applied Biosystems), sequencing reactions were resolved on an ABI 373A automated sequencer.

Results

Searches of the dbEST database (www.ncbi.nih.gov/dbEST) with the BLAST program led to the identification of several overlapping mouse and human sequences that shared high homology to White/ABCG subfamily genes but that appeared to encode a unique gene. Cloning and sequencing identified a cDNA with a single open reading frame encoding 651 amino acids, designated *ABCG5*. This was the longest clone obtained by 5' RACE analysis and the predicted initiation codon matches the consensus sequence. However the open reading frame extends further and we cannot rule out that the protein uses an upstream ATG. Amplification across each of the introns was used to determine that there are 13 exons. The exon size, boundary and splice acceptor and donor sequences and approximate intron sizes are provided in Table 1 below.

Table 1. ABCG5 splice junction sequences.

Exon	Size(bp)	Splice acceptor	Splice donor	Intron (kb)
1	5'		ACAGCGTCAGgtaaggcagagccctt	0.6
2	122	ggggtttcctttaaagCCACCGCGTG	GGAAGCTCAGgtaagcttgggaagga	<6
3	137	tgttgtcgccccgcagGCTCCGGGAA	CGTCCTGCAGgtgggcgcgtccccca	2
4	99	cccgagtctcctgcagAGCGACACCC	CCAGAAGAAGgtgggtgcagcccccc	3
5	133	tttgtgtctcctgcagGTGGAGGCCG	CAGGATCCTAgtaagtggcacccaga	1.4
6	140	ccttctttgctggcagAGGTCATGCT	GCTTTTTCAGgtaagaggttcaactc	1.5
7	130	tctgttgtctggtcagCTCTTTGACA	GACTTCTATAgtaagtttttctttca	0.45
8	214	tgggaaaaacttttagTGGACCTGAC	TTCTCCTGAGgtaagaggctcacaaa	0.1
9	206	ggttgtttgttttcagGAGAGTGACA	GTGAATCTGTgtaagtgcccacgtgc	1
10	139	tgccttccatccccagTTCCCGTGCT	TGTGCTACTGgtgaggggttgttcag	2.5

11	186	gcttatgcttttctagGACGCTGGGC	GATTCCTCAGgtaagatatcataatt	>5
12	113	ttttcttttttcttaagAAACATACAA	TTCACTTGTGgtaagtattctatttg	1.3
13	3'	atcttttccttgacagGCAGCTCAAA		

An amino acid alignment of ABCG5 and several related genes was generated using PILEUP (Genetics Computing Group). After alignment, the sequences were trimmed to minimally overlapping segments and used for neighbor-joining analysis to generate a phylogenetic tree. Fig. 1 displays an alignment of the ABCG5 amino acid sequence with the amino acid sequences of the other ABCG subfamily polypeptides: human ABCG2 (ABCP1), *Drosophila* white (DrWhite), human ABCG1, and the C-terminal half of the yeast YOL075 genes. Identical residues are shaded in black and similar residues in gray. The Walker A, B, and Signature motifs are underlined (A, B, and C) as are the predicted transmembrane segments. Considerable identity is seen in the ATP-binding domain, but there is significant homology throughout the entire coding region.

While all human ABCG genes are half transporters, yeast contains ABCG-type genes that are both half (ADP1) and full transporters (YOL075, PDR5, bfr1). ABCG5 is most closely related to the C-terminal half of the yeast YOL075 gene with 30% overall amino acid identity, 38% identity in the nucleotide binding fold (NBF) and 26% identity in the transmembrane (TM) region. The above-described amino acid sequence alignment was used to generate a phylogenetic tree of the genes, confirming that ABCG5 and YOL075 are closely related. See Fig. 2 where ABCG1 is a human gene; Abcg1 is a mouse gene; YOL075 is a yeast open reading frame (C terminus); bfr1C is a yeast gene (C-terminus); yadp1 is the yeast ADP1, *Drosophila* white gene.

Expression of *ABCG5* in normal human tissues was examined by Northern blot analysis of RNA from human tissues and revealed a 3.5 kb transcript exclusively in the liver. See Fig. 3, where mRNA from brain is in lane 1; 2, from heart; 3, from muscle; 4, from colon; 5, from thymus; 6, from spleen; 7, from kidney; 8, from liver; 9, from intestine; 10, from placenta; 11, from lung; 12, from leukocytes. Real-time PCR analysis showed ABCG5 expression in human intestine, and adult and fetal liver.

Using radiation hybrid analysis, the *ABCG5* gene was mapped to chromosome 2p13-21 between markers *D2S117* and *D2S119*, consistent with data from an *ABCG5* EST (T99836). The mouse *Abcg5* gene was also mapped by radiation hybrids to chromosome 17, 53-55 cM from the centromere. The gene for sitosterolemia, a disorder involving abnormal sterol absorption and defective excretion, also maps to this region.

Example II: The ABCG5 gene is mutated in patients with sitosterolemia

Pedigrees

The pedigrees are shown in Fig. 4. They were recruited based upon previously defined criteria. See Patel *et al. J. Clin. Invest.* 102:1041-1044 (1998) and Patel *et al. J. Lipid. Res.* 39:1055-1061, (1998). Clinical features of some of the probands and their family members have been described previously. Patel *et al. J. Clin. Invest.* 102:1041-1044 (1998) and Patel *et al. J. Lipid. Res.* 39:1055-1061 (1998). Briefly, all probands had clinical features compatible with a diagnosis of sitosterolemia, and all probands had diagnostically elevated plasma sitosterol levels. To date, no other medical condition has been reported to cause elevated plasma sitosterol levels. The pedigrees include six Japanese families (700, 800, 2800, 3300, 3500 and 3700), one South African family of Asian origin (500) and one US Caucasian family (4000). Informed consent was obtained from all participants, in accordance with local Institutional Review Board guidelines.

Exon Amplification and DNA sequencing

Exons were amplified by PCR using oligonucleotide primers located in the flanking intronic area (Table 2). Single-strand conformational polymorphism (SSCP) analyses was performed as previously described. Sossey-Alaoui, *Genomics* 60:330-340 (1999). Direct PCR sequencing was performed using Amplicycle™ Sequencing kit (Perkin Elmer) and analyzed by ABIPRISM™ 377 Genetical Analyzer. Both strands were sequenced to confirm the identified mutations. The primers used for sequencing were the same as those used for PCR amplification. Sequence alignment was aided by the use of MacVector software running on an Apple iMac.

Table 2: Population screening of missense mutations in exon 9 and polymorphisms in exon 13 and primer sequences used for PCR.

	Mutations in Exon 9		Polymorphisms in Exon 13
	Japanese	Caucasian	Caucasian
Heterozygous	0	0	25
Homozygous	0	0	1
Normal	145	156	46
Total	145	156	72

Exon	Forward	Reverse	Product Size, in bp
1	CCCAACTGAAGCCACTCTG	GTGAAGAAAGGCAGCAGA	291
2	GCACAGGTAGGATCAATGCTGG	CAATGTGGAGTTTAACTCAAGCC	267
3	CTCTAGGGCCTTCTGTTG	GCGTCAGTGTAGCCTAAG	232
4	CTTAGGCTACACTGACGC	GGGTGCAAAGGTACTCAG	183
5	CATGTCCTCCCCAGCCCATG	CCAAAGTATCTGCACACACAC	280
6	TGGGCTCTGCACTACCTTAGA	CCTGGCCACTGGTACAAATC	275
7	AAGTGCATCGCTACCCTTGT	GGTGTCATCCAGGCAGAAGT	262
8	CACATGGGTGACATCTTT	TCTCACATTTGTGAGCCT	272
9	GAGGTCTTTAGCCATCCC	AGAAAGAGGTGCACCTCC	308
10	CTAGCCCTCCCTTTTTCAGC	GCAGAGAACTTCACCCTGGA	299
11	ATTCACAGAGGCAAGTGCAG	CCACTATCAGTTCTCTGGTATTCCT	364
12	CTACTGAATTTCATTTTTGTTTTC	CATGCAAAAATAATATCCCCA	184
13	ACACCTTGACACTGTCAA	TTTCCCAGCCATGGCTTT	247

The mutations observed are tabulated in Table 3, below. X denotes a nonsense mutation. The aa numbers in Table 3 indicate the amino acid position in relation to the human ABCG5 cDNA sequence. The presence of the mutation on one or two alleles is indicated. Polymorphic "silent" mutations resulting in no amino acid sequence changes were also observed at the codons for amino acids 9 and 604.

Table 3. ABCG 5 Exon Mutations

Arg243X	Arg243X
Arg419His	Arg419His
Arg389His	Arg389His
del Exon 3	del Exon 3
Arg389His	Arg389His
Arg419His	-
Arg408X	Arg408X
Arg389His	Arg389His
Arg419Pro	Arg419Pro
Glu146Gln	-
Arg408X	-

Based on a sequence-ready BAC contig and transcript map we prepared, we mapped a number of ESTs and genes into the region of interest. Candidate genes were initially screened, based upon whether they were expressed in the liver and/or intestine, the organs important in dietary cholesterol retention. Three ESTs were found to be expressed only in the liver and intestine, one of which, T99836, was found to encode a "half-ABC" transporter, and was studied further. A full-length cDNA was isolated and the gene structure characterized.

The gene consists of 13 exons and encodes a putative six-transmembrane-spanning protein that contains the characteristic ABC signature motif at its N-terminal end. This protein has been assigned the name ABCG5, according to the HUGO nomenclature.

Fig. 4 shows the pedigree of eight sitoserolemia families analyzed for the present study (affected individuals are shown by solid circles or squares, and only the parents are indicated as obligate carriers; carrier status is not shown in unaffected siblings). Probands from the eight families (Fig. 4) were screened using a combination of SSCP analysis and direct sequencing of PCR products. Seven of the probands were expected to carry a homozygous mutation, based upon their haplotype analyses, and one (proband 132) was a potential compound heterozygote.

SSCP analyses indicated potential nucleotide changes in exons 1, 4, 6, 9, and 13. Of these, polymorphic variants in exons 1 and 13 were detected in control samples as well as the probands. Direct sequence analyses showed these to be P9P (exon 1, CCC to CCT) and (exon 13, CAA to GAA). PCR products from probands exhibiting SSCP changes not seen in control DNA, suggestive of mutations, were also directly sequenced.

Fig. 5A-5B shows a composite DNA sequence analysis, as well as the results of a PCR-restriction endonuclease assay of the nucleotide changes identified in the probands, compared with two normal controls. Five mutations, R243Stop (proband 25), R389H (probands 46, 113 and 146), R408Stop (proband 140), R419H (probands 40 and 132) and R419P (proband 157) were identified (Fig. 5A). To confirm that the nucleotide changes were mutations and not polymorphisms, the altered restriction endonuclease recognition sequences were used as an assay. All of the nucleotide changes segregated within the families (see Fig. 5B). Yet another mutation observed was E146Q. Furthermore, screening 82 normal Japanese and 72 US Caucasian individuals and did not identify any carriers for these mutational nucleotide changes.

Polymorphisms, Q604E (exon 13), was identified in many of the probands, as well as the control samples. The probands that were positive for these changes were

heterozygous, rather than homozygous, as expected, based upon their haplotypes. The carrier frequency of Q604E was 35% in the normal US population, with 1% homozygous for this change, suggesting these are polymorphisms. Fig. 6 summarizes the positions of the amino acid changes found in mutant and polymorphic variants of ABCG5.

To exclude that the identified ABCG5 cDNA was a pseudo-gene, all the BACs that define the sitosterolemia locus were screened. Apart from two BACs that are known to span this gene, no other BACs were found to contain this gene. One of the BACs, R489K22, has been sequenced and contains exons 10-13 of ABCG5. BAC R328I4, contains all the exons, based upon PCR data, but has not been sequenced. Thus, gene duplication remains a formal, though remote, possibility. Southern blot analyses of BAC R328I4 with cDNA probes from ABCG5 does not suggest gene duplication.

Example III: Isolation of mouse and rat ABCG5 cDNA

To identify the mouse cDNA, two primers, located in exons 4 and 10, respectively, were used to amplify a fragment from cDNA synthesized from mouse liver. The resultant PCR product was directly sequenced, and a full-length cDNA obtained by 5' and 3' RACE –PCR. The sequence information was used to screen a mouse BAC library to obtain a genomic clone containing exons corresponding to all of the mouse cDNA sequences. A partial rat cDNA clone was identified using the above primers and a rat enterocyte cDNA library as template.

Selectivity for sterol absorption is a feature of other mammals, such as mice, rats, and dogs. Thus, the gene for sitosterolemia would be expected to be highly conserved amongst these species. Isolation of cDNAs encoding the mouse and rat ABCG5 homologues and comparison of their encoded amino acid sequences (Fig. 7) shows that the human and mouse ABCG5 sequences share 85% sequence identity at the amino acid level and 80% at the nucleotide level. The rat sequence, though partial, is also highly conserved. A phylogenetic analysis comparing human, mouse, and rat ABCG5 to other ABC proteins (Fig. 8) shows that the nearest non-mammalian neighbor is a yeast putative ABC protein (YOLO74C, Genbank Accession Nos. Z74816 and Z74817), for which no function has yet been identified. However, a diploid knockout of this gene in yeast is viable, although it exhibits considerable growth delay.

Expression analysis

Expression analyses. Northern-blot analysis was performed as described. Wu *et al. Am. J. Physiol.* 277:E1087-1094 (1999). A multiple-tissue northern blot, containing 2 μg of poly(A)$^+$ RNA (Origene) was hybridized with a full-length mouse cDNA for ABCG5. The hybridized filter was washed stringently with 0.1xSSC/0.1% SDS at 68°C, exposed to a phosphorimager cassette, striped and re-probed with either mouse β-actin or GAPDH probed for comparison of RNA loading. For RT-PCR, human CDNAs (Origene) were used to amplify a fragment spanning exon 1 and 2 using oligonucleotides Wh3fl and Wh3r4. A 250-bp product from cDNA is expected, compared with an 838-bp fragment from the genomic DNA.

Figure 9 shows the results of the Northern Blot. The mRNA was from brain, heart, kidney, liver, lung, muscle, skin, small intestine, spleen, stomach, testis, and the thymus, in lanes 1-12, respectively. As can be seen in Fig. 9, only mRNA from liver and the small intestine hybridized to the mouse cDNA ABCG5. An expected 2.5 kb mRNA was observed, in addition to a fainter band at about 3.3 kb.

The complete sequence analysis of mouse ABCG5 cDNA demonstrated that it encoded an open reading frame of 652 amino acids with a calculated molecular mass of 75 kDa. The deduced amino acid sequence of mouse ABCG5 showed a high degree of conservation, 92.8% and 80.1% matched with rat and human, respectively. Mouse ABCG5 has an extra amino acid, R35, compared to human ABCG5. A poly (A+) site was not identified in the 3' UTR and 3'RACE failed to extend the known 3' end for this cDNA. The ABCG5 protein has a highly conserved ATP-binding cassette signature motif located at the N-terminal half, and a predicted six-transmembrane domain, located at the C-terminal end.

In order to obtain genomic information for both genes, we screened a mouse BAC library (CitbCJ7) using primer sets designed from the first and last exon sequences of mouse *ABCG5*. Exon-intron boundaries were determined by direct sequencing of the BAC DNA and/or long PCR amplified products using exon specific primers. All exon-intron boundaries show canonical sequences with initial GT as splice donor and terminal AG as splice acceptor followed the rule of splice junctions.

A single TATA box was identified 232 bp up-stream of mouse *ABCG5* initiator codon, as well as a GATA motif, and the analyses predicted a potential 'promoter' site. This promoter region sequence has a 40% homology to the human sequence. There are two regions that show a very high degree of conservation between human and mouse sequences, although the human sequence does not contain an identifiable TATA or CCAT motif. Berge *et al* identified ABCG5 cDNAs as transcripts that were induced after rexinoid exposure,

suggesting that LXR-RXR may be involved in their regulation. Berge *et al. Science* 290:1771-1775 (2000). Repa *et al.* showed that LXR deficiency affected cholesterol absorption. Repa *et al. Science* 289:1524-9 (2000). Thus LXR is a strong candidate as a regulatory transcriptional factor.

Genetic variations in in-bred mouse strains

In-bred mouse strains have been used to identify genes whose genetic variations may be important determinants of arteriosclerosis, gall stone formation or biliary cholesterol secretion. Nishina *et al. Lipids* 28:599-605 (1993); Purcell-Huynh *et al. J. Clin. Invest.* 96:1845-58 (1995); Mehrabian *et al. J. Lipid Res.* 41:1936-46 (2000); Paigen *et al. Physiol. Genomics.* 4:59-65 (2000); Perusse *et al. Obesity Res.* 9:135-69 (2001). Some of these in-bred mouse strains have been screened for differences in dietary cholesterol absorption. Kirk *et al. J. Lipid Res.* 36:1522-32 (1995); Howles *et al. J. Biol. Chem.* 271:7196-202 (1996); Carter *et al. J. Nutr.* 127:1344-1348 (1997); Jolley *et al. Am. J. Physiol.* 276:G1117-G1124 (1999).

To identify whether genetic variations in *ABCG5* may be responsible for some of these phenotypes, 17 strains were screened. These strains were selected, based upon either documentation of cholesterol absorption rates, or having very high levels of plasma cholesterol levels. The latter phenotype was chosen because in some sitosterolemia patients presented with very high levels of plasma cholesterol and were initially diagnosed as pseudohomozygous familial hypercholesterolemia. Both coding and non-coding alterations were detected for *abgc5*, including polymorphisms that altered amino acid coding and single nucleotide changes in exonic regions that did not alter amino acid coding. All of these changes were present as homozygous changes, compatible with the in breeding of these lines.

SEQUENCES

SEQ. ID NO. 40

Human ABCG5 polypeptide sequence (Genbank AF312715)

MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWWD
ITSCRQQWTRQILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLGEV
YVNGRALRREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRRGNPGSFQKKVE
AVMAELSLSHVADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDC

MTANQIVVLLVELARRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFF
NDCGYPCPEHSNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKSAICHKTLKN
IERMKHLKTLPMVPFKTKDSPGVFSKLGVLLRRVTRNLVRNKLAVITRLLQNLIMG
LFLLFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFPVLRAVSDQES
QDGLYQKWQMMLAYALHVLPFSVVATMIFSSVCYWTLGLHPEVARFGYFSAALL
APHLIGEFLTLVLLGIVQNPNIVNSVVALLSIAGVLVGSGFLRNIQEMPIPFKIISYFTF
QKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMCAFTQGIQFIEKTCPGATSRFTMNF
LILYSFIPALVILGIVVFKIRDHLISR*

SEQ. ID NO. 41

Human ABCG5 coding sequence

ATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCTGGAGGGGGCTCCTGCC
ACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCCTACAGCGTCAGC
CACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGACCAG
GCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGTGCAT
CCTAGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGACGCCATGTCCGGGA
GGCTGGGGCGCGCGGGGACCTTCCTGGGGGAGGTGTATGTGAACGGCCGGGCG
CTGCGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGTCCTGCAGAGCGACACC
CTGCTGAGCAGCCTCACCGTGCGCGAGACGCTGCACTACACCGCGCTGCTGGCC
ATCCGCCGCGGCAATCCCGGCTCCTTCCAGAAGAAGGTGGAGGCCGTCATGGCA
GAGCTGAGTCTGAGCCATGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGG
GGCATTTCCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGCCCAGCTGCTCCAG
GATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGCCTGGACTGCATGACT
GCTAATCAGATTGTCGTCCTCCTGGTGGAACTGGCTCGCAGGAACCGAATTGTG
GTTCTCACCATTCACCAGCCCCGTTCTGAGCTTTTTCAGCTCTTTGACAAAATTG
CCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGCGGAAATGCTTG
ATTTCTTCAATGACTGCGGTTACCCTTGTCCTGAACATTCAAACCCTTTTGACTT
CTATATGGACCTGACGTCAGTGGATACCCAAAGCAAGGAACGGGAAATAGAAA
CCTCCAAGAGAGTCCAGATGATAGAATCTGCCTACAAGAAATCAGCAATTTGTC
ATAAAACTTTGAAGAATATTGAAAGAATGAAACACCTGAAAACGTTACCAATG
GTTCCTTTCAAAACCAAAGATTCTCCTGGAGTTTTCTCTAAACTGGGTGTTCTCC
TGAGGAGAGTGACAAGAAACTTGGTGAGAAATAAGCTGGCAGTGATTACGCGT
CTCCTTCAGAATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGTTCTGCGGGTCCG
AAGCAATGTGCTAAAGGGTGCTATCCAGGACCGCGTAGGTCTCCTTTACCAGTT

TGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGT
GCTGCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTACCAGAAGTGGC
AGATGATGCTGGCCTATGCACTGCACGTCCTCCCCTTCAGCGTTGTTGCCACCAT
GATTTTCAGCAGTGTGTGCTACTGGACGCTGGGCTTACATCCTGAGGTTGCCCG
ATTTGGATATTTTTCTGCTGCTCTCTTGGCCCCCCACTTAATTGGTGAATTTCTAA
CTCTTGTGCTACTTGGTATCGTCCAAAATCCAAATATAGTCAACAGTGTAGTGG
CTCTGCTGTCCATTGCGGGGGTGCTTGTTGGATCTGGATTCCTCAGAAACATACA
AGAAATGCCCATTCCTTTTAAAATCATCAGTTATTTTACATTCCAAAAATATTGC
AGTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTTCACTTGTGGCAGCT
CAAATGTTTCTGTGACAACTAATCCAATGTGTGCCTTCACTCAAGGAATTCAATT
CATTGAGAAAACCTGCCCAGGTGCAACATCTAGATTCACAATGAACTTTCTGAT
TTTGTATTCATTTATTCCAGCTCTTGTCATCCTAGGAATAGTTGTTTTCAAAATA
AGGGATCATCTCATTAGCAGGTAG

SEQ. ID NO. 42

Human ABCG5 cDNA sequence

AAGTCCCAGTCCTGCTGTCCCAAGGGACTCCGGGGTCAGGTGGAGCAGGCAGG
GCAGTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGAGGGTCCGGCCACC
AGAAAATTTGCCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTCATCTTTGA
CCCCCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCTGG
AGGGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCT
CCTACAGCGTCAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGC
AGCAGTGGACCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGG
CAGATCATGTGCATCCTAGGAAGCTCAGGCTCCGGGAAAACCACGCTGCTGGAC
GCCATGTCCGGGAGGCTGGGGCGCGCGGGGACCTTCCTGGGGGGAGGTGTATGT
GAACGGCCGGGCGCTGCGCCGGGAGCAGTTCCAGGACTGCTTCTCCTACGTCCT
GCAGAGCGACACCCTGCTGAGCAGCCTCACCGTGCGCGAGACGCTGCACTACA
CCGCGCTGCTGGCCATCCGCCGCGGCAATCCCGGCTCCTTCCAGAAGAAGGTGG
AGGCCGTCATGGCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTGATTGGC
AACTACAGCTTGGGGGGCATTTCCACGGGTGAGCGGCGCCGGGTCTCCATCGCA
GCCCAGCTGCTCCAGGATCCTAAGGTCATGCTGTTTGATGAGCCAACCACAGGC
CTGGACTGCATGACTGCTAATCAGATTGTCGTCCTCCTGGTGGAACTGGCTCGC
AGGAACCGAATTGTGGTTCTCACCATTCACCAGCCCCGTTCTGAGCTTTTTCAGC
TCTTTGACAAAATTGCCATCCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCC

AGCGGAAATGCTTGATTTCTTCAATGACTGCGGTTACCCTTGTCCTGAACATTCA
AACCCTTTTGACTTCTATATGGACCTGACGTCAGTGGATACCCAAAGCAAGGAA
CGGGAAATAGAAACCTCCAAGAGAGTCCAGATGATAGAATCTGCCTACAAGAA
ATCAGCAATTTGTCATAAAACTTTGAAGAATATTGAAAGAATGAAACACCTGAA
AACGTTACCAATGGTTCCTTTCAAAACCAAAGATTCTCCTGGAGTTTTCTCTAAA
CTGGGTGTTCTCCTGAGGAGAGTTACAAGAAACTTGGTGAGAAATAAGCTGGCA
GTGATTACGCGTCTCCTTCAGAATCTGATCATGGGTTTGTTCCTCCTTTTCTTCGT
TCTGCGGGTCCGAAGCAATGTGCTAAAGGGTGCTATCCAGGACCGCGTAGGTCT
CCTTTACCAGTTTGTGGGCGCCACCCCGTACACAGGCATGCTGAACGCTGTGAA
TCTGTTTCCCGTGCTGCGAGCTGTCAGCGACCAGGAGAGTCAGGACGGCCTCTA
CCAGAAGTGGCAGATGATGCTGGCCTATGCACTGCACGTCCTCCCCTTCAGCGT
TGTTGCCACCATGATTTTCAGCAGTGTGTGCTACTGGACGCTGGGCTTACATCCT
GAGGTTGCCCGATTTGGATATTTTTCTGCTGCTCTCTTGGCCCCCCACTTAATTG
GTGAATTTCTAACTCTTGTGCTACTTGGTATCGTCCAAAATCCAAATATAGTCAA
CAGTGTAGTGGCTCTGCTGTCCATTGCGGGGGTGCTTGTTGGATCTGGATTCCTC
AGAAACATACAAGAAATGCCCATTCCTTTTAAAATCATCAGTTATTTTACATTCC
AAAAATATTGCAGTGAGATTCTTGTAGTCAATGAGTTCTACGGACTGAATTTCA
CTTGTGGCAGCTCAAATGTTTCTGTGACAACTAATCCAATGTGTGCCTTCACTCA
AGGAATTCAATTCATTGAGAAAACCTGCCCAGGTGCAACATCTAGATTCACAAT
GAACTTTCTGATTTTGTATTCATTTATTCCAGCTCTTGTCATCCTAGGAATAGTTG
TTTTCAAAATAAGGGATCATCTCATTAGCAGGTAGTGAAAGCCATGGCTGGGAA
AATGGAAGTGAAGCTGCCGACTGTGCATGACTGCTCTGAACGTCTGAAATGAGA
GTGCCATGTATTTCTTTCTTGACAGGACATCTCAAGTCTTTTAACCATTAAGACT
CCATTTGTGCCTCTTGGATCCAAGCAGGCCTTGAATGCAATGGAAGTGGTTTAT
AGTCCCTTGCTCTTACAACTTGCAGGGACATGTGGTTATTTGGAAATTGTGACTG
AGCGGACCCAAGAATGTAAATAATATTCATAAACCTATGGGAGACTCGTGTGAC
TATTTTTTTTCCTTGTTCTAGGCACAGAAAAAAATAGGTCAGCTTAAAAATATGT
TTACATTGGATAAAGGATTAGGCAAAAATAAAATGTTTCAAGGATTCCTGACCA
TAAGTGACAGAGAAAGAGAG

SEQ. ID NO. 43

Sequence of human ABCG5 upstream genomic sequence, exon 1, intron 1, and exon 2

GCTTAGATTTTGGCAGATGAGACAGGTGATTACGATGGAGCGAGACAGACCAG
GCAGAGAGGGAGGCTGACCTCAAGCATCTGACCAGGGTTTATTTTCACTATGTA

CACAATGCACGCACATGCTGTAGGTGAAACTAACATGACTTATGCCTTACGTGA
AAATAACAAATAACATACAGCAGTCTTCTTGTCAAAGTACCCCTGCAATGATGG
GGGCCAGAAGTTCTCGGAAAGAACATTCCAGGTCAGTGGAGGTGGAGGGAAAG
AAATTCGGTGACAGTATGCCGCAGGCGTGCTGTGGGGAACCCTTAGAGTTCTGG
GAGAGTATGAAAAACAGCAGACGAAAGTGACTTCCATTGCTTAATGTTTGAATT
ATCACCATAAAGACCCAAAATTATACAGAAAAAAATGGGAAAGATAAACACCT
AATTCAGGAGAGGGGTTATCTCTGGGAGCGGGGAGAGATGAAAGGAGAAAGG
GACACAGTAGGAGTGCGGGGATCAAGCTTNAAAGCTTTTGGTAATTGGTGGTGT
GTACATAGAGTTCTTTAAGCTATTTGTGTCTTGGTAATTGGTGGTGTGTACATGA
GAGTTCTTTAAGCTGATTTGTGTACGATATTTCACAATATGTTGTCAAATTGAGA
GAGCAAGCCAGTGAGTAGAACTCCAGAGTTCCATTCCCACCTCAACCCCAGTTG
CCAAAGCCCCCAAGCAGGAGGGAGGTTGAGGGAGAGGAGGTAAGAAGGTCTGT
GCCCCAAGCTCCTGAGGAGATTGAAAGCAGCTCGGACAGATGCTGGACTCCTG
GCAGATCAGCCCTTCGGCCTTGCCCCTCACTCGCCCTCTCCGCTGTCACTGTGCA
CATTGCTTCATTGTCCCATTTTTGTTGTTGTTGTTGAATCATCAAAAAATCTTAGC
CATTGCCAACTGTGCGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAAGC
TGAGGAGGGCAGATCACCTGAGGCCAGGAGTTCAAGACCAGCTTGGCCAAAAC
AGTGAAACCCTGTCTCTACTAAAAATACAAAAAAAAAATTGCCGGGCATGGTGG
CACACCCCTATAGTCCCAGCTACCCTACTCAAGAGGTTGAGGCAGGAGAATCAC
TTGAACCGGCAGGTGGAGGTGCAGGGAGCCGAGATCGTGCCTCTGCACTCCAG
CCTGGGCCACAGCGTGAGACTCTGTCAAAAAAAAAAAAAATCTTATCCATTTCTA
AGAGACTCATGTGGGCTAACATGCATCTTGCTGTTGTTGTTTTTAAAACAAATAT
CTGCAGGAGGGATATTAGACAATGTAAATGAGCTTGAGAAATTACTTCTGCTGG
CCAAAACATCCCCAGGTCGGATTTAGGCATGGAGGAGGCGCTGAGGAAGGAAG
GTCAGACATTTGGAGTCTGGAGCAATGTGTGGAGGTAACCTGCAGCCCAACTGG
GTTCCACTGTGTGGTGCTTTGAGAAGGGGAGGGAGCTGGCAGACAGAGGAGGC
AGAGCAGAAACTGGGTAAGAGGAAGGGGAGAGGTGCCTGGTTGCTTGTTCCAA
GGCTCACCCCAAGCCCCTTCACTGTGGCCTGGGAAGGCAGGTGTGGGCCAGCCC
TGACTCCAACCACCATTGAGGGATTGTGCCTTCCAGGAGTTGCACAGGCAGCTC
ATCTTTCGGAGGAAGAAAATTCGTGCCTGGGTTGGGGACGACCTCTGTTCCATT
TTAAAATATTTTCCTTGGCTCCCAGGAAGGATTTGTTAGACTCTTCCTGAGGTTT
TGACAAACTCTCTGTATTTTTCAAATACTTAAGTATCTATTCGGCTGACATCTTA
ATCAGTACGACTGTCAGAACATCACTTGAATTTCTGACAGGTGACACCCAAAAA
AGCAAAAAGCAGGTTTATTTGTAGGTAACCAGCTCTGCTCATGCTGGGGCTACA

TTGTAATTTCTCCTCGTATTAACTTCTGATCAAATTCCTGAGTCAGATGCCTAGG
CAAGAAGGAAACTCACAGAGCACATGTTTCTAGTTCTGAGATGAGGAGCCTATG
CCCCGGGGGGAGTGATGTGCTGACACTCACGGCTGGAGGGTTGGCAAGAGGAC
ACGCAGGACTTGTTCCTGGCTGAAGAAATTTTATCGAAACATTCAGCCTAGGTC
ACACACAGCTCTGCCTGCTGCCAGGGTTTCTCTTGTCCTTCTCTGTTGCTGCTCCT
GCCCATGGCATGAGGAGTTTGTGGGTTAAGGGCACTTGCCACTCCAGGTGCCCA
AGATGCCAGATATTCTCTGTGCAAATGGCCCCAAGTCCATCCCCAGGGTCTGTA
CACCTCTTCCCAGGCCCAACCTCCTGAGGACTTTTAGGCCAGAGAAGTATATGT
CTGTGCAGGGCAGGGCTGCAATGCAGGTAGGCAGGAGGGTGACCATCCAGGGT
GTTGAGGGCCCCATGGGAGTGGGATGGAGCTGAGGGGGCCCAGAAAAGGGGG
GACAGGGGTTGTGGGCTCGGGGCTGGGAAGGTGGCTTCCCCCGTACCAGCCGC
ATTCTAAGCCCAAGGTGGCCNTAAGAAATTTCTTCAAATTTACACATGGGCCCT
TTCAGGTTGGTGGAGGGAAGAATATGGTCAAGGATAAGGANGGATAGGAAACT
ATTTTAATTTANACTGGGTCTTATAAGCTTTGGACTGGATGGACTTTTATATATC
TAAGACATAGGGAAATGGTGACCCTCATCCTCTTGGTTCCAAGACCCACAAGGT
GTTACGGGNCAGGACCCGTCACTCAAGCACCTGGAGTGACAGGGATCCGGGAC
AAGAAGGAAGCAGAAATGGCAGGGCCTGCGTGCATTTCTGGTGTGGTCCTAGC
ATTTCTTTGCCTCTCAAGCTGTGGTGACTGAATCGTCAGCCCTCCCAGGCAGAA
GTGTTCCCAAAGTCCCGGATGACTTTCTATTCTATTCAGGCTTTAAACATTTCCC
GAAGTGGTGGGCCCCACAGGGTATTCANAGAGCAGAGCTGGTCAGATGTGGTG
GTTGCAGAACTGACTAGAAATGGTGGGCTCCTTGGGTCTGACCGAGTCAAGTCC
TGAAACTCAAGGCCAGTCCAGGTTGTTTTCCCCATTGGGTGTGGAATCCTCGAT
CATGCATGTCTTTCTCTCCTCCCTCCTACCCACAACAGGCAAAGATGGAAAGGT
AGAATGGGGTGAGGTGGTGGGAGTGGGGATCTGCTTCTCGCTTGTCTTCCAGTT
TAGCCTCGTGCTTCAAATCCTGCCACATCCCGAATTCAGTCAAAGGCTATTTCTT
GAGTAAACACTTCTCAGGTAAAATGAGGAAGGAAACATACCTCCACCTCCTGCC
ACTTGGCTGCTTCTACTCCTTCCAGCTTTCTCGCAGAACTTACGATTGCCTGTTA
GAGCCACACATGCTGATGTTCCCACAAAAGCCGTCTGTCACCCTGTCTCACCCA
TCACTTCATTACTGAGACCAACCTGACAGCCACATGGATGGAGGAGACACCAGT
GGAGTCCCTGGCATGTGTGAGCACATTTCTATGATGGAGTCTCATTCGGAAAAA
GCGAAACTGGCCAGGCACGGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGG
CCGAGGAGGGTGGGTAACCTGAGGTCAGGAGTTTGAGCCTGACCAACATGGTG
AAACACTGTCTCTACTAAAAAATACAAAATTTGCCAGGTGTGGTGGCAGGTGCC
TGTAATCTCAGCTACTTGGGAGGCTGAGGCAGGAGAATCACTTGAACCTAGGAG

GCAGAGCTTGCAGTGAGCCGAGATCACGCCACTGCACTCCAGCTTGGGCAACA
GAGCGAGACTCCGTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AANANGGGGAAGAAAAGAAAAAAGAAAAAGTGAAATTGTCCCACATCACACA
AAAGAACATCATTTCCCTAAAAGAGCATTTCTTAGGGCAGGAAGTGACCTCAGA
GGCCTCTGGGACCCTGAATCTGTTCCCCTCCGCCCTTTGACATGCAGGAAACAG
TCCTGCGGCCATGTCCTCACACTGCTTGATGTCCGGGTGGTGCTAGGACAGAAG
GCTCCTGAGGGAAGAGAGAAAGGTTTGATTTCTCCTACCCGCCCACCAGGCCTG
GGCCGACTTCCCATTGCTCACTCACCGAGGTATCCTGGGGAGTGGCCCCTTTCG
GCAGCCCTCTCTCCTCTGCCGCCTTCCCGGCCATGGGGCCCACAGGTCTGTGACC
CTGGGCTGCAGCTCTCTTAGACCCAGCTGCTGCCTGCCAGGGCCAGTGTCTTCA
CTCTGTTTCTTGGAGCAGGGACACCTCGGCCTCCTGCCCTGGGCCCGTCTCTCCC
AGCATTCCTTGCTGGCAAGCCCACCTACAAACGTGTGTGTTCTTGCCCACTGTCA
AGATAAGGACGCGCTGGCTAAAGGTACATCAGATAATGGTCTCCGTGGCCAAG
TCCCAGTCCTGCTGTCCCAAGGGACTCCGGGGTCAGGTGGAGCAGGCAGGGCA
GTCTGCCACGGGCTCCCCAACTGAAGCCACTCTGGGGAGGGTCCGGCCACCAGA
AAATTTGCCCAGCTTTGCTGCCTGTTGGCCATGGGTGACCTCTCATCTTTGACCC
CCGGAGGGTCCATGGGTCTCCAAGTAAACAGAGGCTCCCAGAGCTCCCTGGAG
GGGGCTCCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCATCCTCCATGCCTCC
TACAGCGTCAGGTAAGGCAGAGCCCTTGCTGCTGCTGCTCCCCCAGGAGTGCGG
GGCCCGGCGCTCACCCCTCTGCTGCCTTTCTTCACTCTTTAAGTGCCAGTCTGGG
CACTTCGGGCTCCCTCTTTAGTGGATCGGGTGGAGAGAGGAGAGGGAGAAGGG
CTGTTGCTGGGAAACATGGAGCGACAGTGAATGGCCCCTCCCCCTGCCCAGGGA
AGGGCCTGGGCATAAACAAAGTGGCAGCAGTGCCCTGCCAACCCAGTGTCTAC
GGCCTGCCCTCTGTGGATGGGAATGGGGGTACTGCGAATGCAAGGAGTCTTGAA
ACCTGGTGAAAGAATGCAGGGACAGCCACCTCGCAGCCAAACGGACAGGACAT
TCAGAGCAACTCCAGCACAGGCCCCCTCCCTACGTGGCAGACAGCCTCAGTCGC
TATCTGCCAGGTTCTACAGAGGAGGGCGCAGAGACTGAAACACGTTAGGAGCC
TGTCCGGAGACTACTGGGGGTGGGGCACAGGTAGGATCAATGCTGGGGACCTG
GGTGTGGCCCCTTCCAGGGCCCCAAGCTGCCTTTGCCTTCCTGGGGTTTCCTTTA
AAGCCACCGCGTGAGGCCCTGGTGGGACATCACATCTTGCCGGCAGCAGTGGA
CCAGGCAGATCCTCAAAGATGTCTCCTTGTACGTGGAGAGCGGGCAGATCATGT
GCATCCTAGGAAGCTCAGGTA

SEQ. ID NO. 44

Mouse ABCG5 polypeptide sequence (Genbank AF312713)

MGELPFLSPEGARGPHINRGSLSSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWW
NIKSCQQKWDRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRLRRTGTLEGEV
FVNGCELRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNKKVE
AVMTELSLSHVADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDC
MTANQIVLLLAELARRDRIVIVTIHQPRSELFQHFDKIAILTYGELVFCGTPEEMLGFF
NNCGYPCPEHSNPFDFYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILEN
IERARYLKTLPTVPFKTKDPPGMFGKLGVLLRRVTRNLMRNKQAVIMRLVQNLIMG
LFLIFYLLRVQNNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMLRAVSDQE
SQDGLYHKWQMLLAYVLHVLPFSVIATVIFSSVCYWTLGLYPEVARFGYFSAALLA
PHLIGEFLTLVLLGIVQNPNIVNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQK
YCCEILVVNEFYGLNFTCGGSNTSMLNHPMCAITQGVQFIEKTCPGATSRFTANFLIL
YGFIPALVILGIVIFKVRDYLISR*

SEQ. ID NO. 45

Mouse ABCG5 coding sequence

ATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCACATCAAC
AGAGGGTCTCTGAGCTCCCTGGAGCAAGGTTCGGTCACGGGCACAGAGGCTCG
GCACAGCTTAGGTGTCCTGCATGTGTCCTACAGCGTCAGCAACCGTGTCGGGCC
TTGGTGGAACATCAAATCATGCCAGCAGAAGTGGGACAGGCAAATCCTCAAAG
ATGTCTCCTTGTACATCGAGAGTGGCCAGATTATGTGCATCTTAGGCAGCTCAG
GCTCAGGGAAGACCACGCTGCTGGACGCCATCTCCGGGAGGCTGCGGCGCACT
GGGACCCTGGAAGGGGAGGTGTTTGTGAATGGCTGCGAGCTGCGCAGGGACCA
GTTCCAAGACTGCTTCTCCTACGTCCTGCAGAGCGACGTTTTTCTGAGCAGCCTC
ACTGTGCGCGAGACGTTGCGATACACAGCGATGCTGGCCCTCTGCCGCAGCTCC
GCGGACTTCTACAACAAGAAGGTAGAGGCAGTCATGACAGAGCTGAGCCTGAG
CCACGTGGCGGACCAAATGATTGGCAGCTATAATTTTGGGGGAATTTCCAGTGG
CGAGCGGCGCCGAGTTTCCATCGCAGCCCAACTCCTTCAGGACCCCAAGGTCAT
GATGCTAGATGAGCCAACCACAGGACTGGACTGCATGACTGCAAATCAAATTGT
CCTTCTCTTGGCTGAGCTGGCTCGCAGGGACCGAATTGTGATTGTCACCATCCAC
CAGCCTCGCTCTGAGCTCTTCCAACACTTCGACAAAATTGCCATCCTGACTTACG
GAGAGTTGGTGTTCTGTGGCACCCCAGAGGAGATGCTTGGCTTCTTCAATAACT

GTGGTTACCCCTGTCCTGAACATTCCAATCCCTTTGATTTTTACATGGACTTGAC
ATCAGTGGACACCCAAAGCAGAGAGCGGGAAATAGAAACGTACAAGCGAGTAC
AGATGCTGGAATGTGCCTTCAAGGAATCTGACATCTATCACAAAATTCTGGAGA
ACATTGAAAGAGCACGATACCTGAAAACCTTACCCACGGTTCCTTTCAAAACAA
AAGATCCTCCTGGGATGTTCGGCAAGCTTGGTGTCCTGCTGAGGCGAGTAACAA
GAAACTTAATGAGGAATAAGCAGGCAGTGATTATGCGTCTCGTTCAGAATCTGA
TCATGGGCCTCTTCCTCATTTTCTACCTTCTCCGCGTCCAGAACAACACGCTAAA
GGGCGCTGTGCAGGACCGCGTGGGGCTGCTCTATCAGCTTGTGGGTGCCACCCC
ATACACCGGCATGCTCAATGCTGTGAATCTGTTTCCCATGCTGAGAGCCGTCAG
CGACCAGGAGAGTCAGGATGGCCTGTATCATAAGTGGCAGATGCTGCTCGCCTA
CGTGCTACACGTCCTCCCCTTCAGCGTCATCGCCACGGTCATTTTCAGCAGTGTG
TGTTATTGGACTCTGGGCTTGTATCCTGAAGTTGCCAGATTTGGATATTTCTCTG
CTGCTCTTTTGGCCCCTCACTTAATTGGAGAATTTCTAACACTTGTGCTGCTTGG
TATAGTCCAAAACCCTAATATTGTCAACAGTATAGTGGCTCTGCTCAGCATCTCT
GGGCTGCTTATTGGATCTGGATTTATCAGAAACATACAAGAAATGCCCATTCCT
TTAAAAATCCTGGGTTATTTTACATTCCAAAAATACTGTTGTGAGATTCTCGTGG
TCAATGAGTTTTACGGCCTGAACTTCACTTGTGGTGGATCCAACACCTCTATGCT
AAATCACCCGATGTGCGCCATCACCCAAGGGGTCCAGTTCATCGAGAAAACCTG
CCCAGGTGCTACATCCAGATTCACGGCAAACTTCCTCATCTTATATGGGTTTATC
CCAGCTCTGGTCATCCTAG

SEQ. ID NO. 46

Mouse ABCG5 cDNA sequence (Genbank AF312713)

ATTGGTGAACTGTTATCTCACGAGGATTCCAGGGCTGGGTAGGATCGGACAGGG
CACTCCCATTGGCTCCTCAGTTAAAGCTGCCCTGGAGCCGGACAGGCCACTAGA
AAATTCACTTGCATTTGCTTCCTGCTAGCCATGGGTGAGCTGCCCTTTCTGAGTC
CAGAGGGAGCCAGAGGGCCTCACATCAACAGAGGGTCTCTGAGCTCCCTGGAG
CAAGGTTCGGTCACGGGCACAGAGGCTCGGCACAGCTTAGGTGTCCTGCATGTG
TCCTACAGCGTCAGCAACCGTGTCGGGCCTTGGTGGAACATCAAATCATGCCAG
CAGAAGTGGGACAGGCAAATCCTCAAAGATGTCTCCTTGTACATCGAGAGTGGC
CAGATTATGTGCATCTTAGGCAGCTCAGGCTCAGGGAAGACCACGCTGCTGGAC
GCCATCTCCGGGAGGCTGCGGCGCACTGGGACCCTGGAAGGGGAGGTGTTTGT
GAATGGCTGCGAGCTGCGCAGGGACCAGTTCCAAGACTGCTTCTCCTACGTCCT
GCAGAGCGACGTTTTTCTGAGCAGCCTCACTGTGCGCGAGACGTTGCGATACAC

AGCGATGCTGGCCCTCTGCCGCAGCTCCGCGGACTTCTACAACAAGAAGGTAGA
GGCAGTCATGACAGAGCTGAGCCTGAGCCACGTGGCGGACCAAATGATTGGCA
GCTATAATTTTGGGGGAATTTCCAGTGGCGAGCGGCGCCGAGTTTCCATCGCAG
CCCAACTCCTTCAGGACCCCAAGGTCATGATGCTAGATGAGCCAACCACAGGAC
TGGACTGCATGACTGCAAATCAAATTGTCCTTCTCTTGGCTGAGCTGGCTCGCA
GGGACCGAATTGTGATTGTCACCATCCACCAGCCTCGCTCTGAGCTCTTCCAAC
ACTTCGACAAAATTGCCATCCTGACTTACGGAGAGTTGGTGTTCTGTGGCACCC
CAGAGGAGATGCTTGGCTTCTTCAATAACTGTGGTTACCCCTGTCCTGAACATTC
CAATCCCTTTGATTTTTACATGGACTTGACATCAGTGGACACCCAAAGCAGAGA
GCGGGAAATAGAAACGTACAAGCGAGTACAGATGCTGGAATGTGCCTTCAAGG
AATCTGACATCTATCACAAAATTCTGGAGAACATTGAAAGAGCACGATACCTGA
AAACCTTACCCACGGTTCCTTTCAAAACAAAAGATCCTCCTGGGATGTTCGGCA
AGCTTGGTGTCCTGCTGAGGCGAGTAACAAGAAACTTAATGAGGAATAAGCAG
GCAGTGATTATGCGTCTCGTTCAGAATCTGATCATGGGCCTCTTCCTCATTTTCT
ACCTTCTCCGCGTCCAGAACAACACGCTAAAGGGCGCTGTGCAGGACCGCGTGG
GGCTGCTCTATCAGCTTGTGGGTGCCACCCCATACACCGGCATGCTCAATGCTG
TGAATCTGTTTCCCATGCTGAGAGCCGTCAGCGACCAGGAGAGTCAGGATGGCC
TGTATCATAAGTGGCAGATGCTGCTCGCCTACGTGCTACACGTCCTCCCCTTCAG
CGTCATCGCCACGGTCATTTTCAGCAGTGTGTGTTATTGGACTCTGGGCTTGTAT
CCTGAAGTTGCCAGATTTGGATATTTCTCTGCTGCTCTTTTGGCCCCTCACTTAA
TTGGAGAATTTCTAACACTTGTGCTGCTTGGTATAGTCCAAAACCCTAATATTGT
CAACAGTATAGTGGCTCTGCTCAGCATCTCTGGGCTGCTTATTGGATCTGGATTT
ATCAGAAACATACAAGAAATGCCCATTCCTTTAAAAATCCTGGGTTATTTTACA
TTCCAAAAATACTGTTGTGAGATTCTCGTGGTCAATGAGTTTTACGGCCTGAACT
TCACTTGTGGTGGATCCAACACCTCTATGCTAAATCACCCGATGTGCGCCATCA
CCCAAGGGGTCCAGTTCATCGAGAAAACCTGCCCAGGTGCTACATCCAGATTCA
CGGCAAACTTCCTCATCTTATATGGGTTTATCCCAGCTCTGGTCATCCTAGGAAT
AGTGATTTTTAAAGTCAGGGACTACCTGATTAGCAGATAGTTAAGATGACAGGC
AGGAAAGGGTTAATGGGCAGGCACGCCCACTGTGGAGCACAGAGAAGTACTGT
CTACAACCATCAGGATTCCATCTGCGACCCTTGTGTCTGACCCTTGTGTCTATCC
GGAGCCCCAAGGGCAACGAGAACTCACAGCCCTCTGCTATTCCAGCTTGTGGGG
CAATGTGGTGCTTGGACATTGTGACTGAACTGGTCCAATAATGTAAATAATAAT
AATTCATAAACCTACAGGACATTAAAA

SEQ. ID NO. 47

Rat AGCG5 polypeptide sequence (Genbank AF312714)

MGELPFLSPEGARGPHNNRGSQSSLEEGSVTGSEARHSLGVLNVSFSVSNRVGPWW
NIKSCQQKWDRKILKDVSLYIESGQTMCILGSSGSGKTTLLDAISGRLRRTGTLEGE
VFVNGCELRRDQFQDCVSYLLQSDVFLSSLTVRETLRYTAMLALRSSSADFYDKKV
EAVLTELSLSHVADQMIGNYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLD
CMTANHIVLLLVELARRNRIVIVTIHQPRSELFHHFDKIAILTYGELVFCGTPEEMLG
FFNNCGYPCPEHSNPFDFYMDLTSVDTQSREREIETYKRVQMLESAFRQSDICHKIL
ENIERTRHLKTLPMVPFKTKNPPGMFCKLGVLLRRVTRNLMRNKQVVIMRLVQNLI
MGLFLIFYLLRVQNNMLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMLRAVS
DQESQDGLYQKWQMLLAYVLHALPFSIVATVIFSSVCYWTLGLYPEVARFGYFSAA
LLAPHLIGEFLTLVLLGMVQNPNIVNSIVALLSISGLLIGSGFIRNIEEMPIPLKILGYFT
FQKYCCEILVVNEFYGLNFTCGGSNTSVPNNPMCSMTQGIQFIEKTCPGATSRFTTN
FLILYSFIPTLVILGMVVFKVRDYLISR*

SEQ. ID NO. 48

Rat AGCG5 cDNA (Genbank AF312714)

GCTGGCCATGGGTGAGCTGCCCTTTCTGAGTCCAGAGGGAGCCAGAGGGCCTCA
CAACAACAGAGGGTCTCAGAGCTCCCTGGAGGAAGGCTCAGTTACAGGCTCAG
AGGCTCGGCACAGCTTAGGTGTCCTGAATGTGTCCTTCAGCGTCAGCAACCGTG
TCGGGCCCTGGTGGAACATCAAATCATGCCAGCAGAAGTGGGACAGGAAAATC
CTCAAAGATGTCTCCTTGTACATCGAGAGTGGCCAGACCATGTGCATCTTAGGT
AGCTCAGGCTCAGGGAAAACCACGCTGCTGGACGCCATCTCTGGGAGGCTGCG
GCGCACAGGGACCTTGGAAGGGGAAGTGTTTGTGAACGGCTGCGAGCTGCGCA
GGGACCAGTTCCAAGACTGCGTCTCCTACCTCCTGCAGAGCGATGTCTTTCTGA
GCAGCCTCACGGTGCGGGAGACGCTGAGATACACGGCGATGCTGGCTCTCCGC
AGCAGCTCCGCGGACTTCTACGACAAGAAGGTAGAGGCAGTCCTGACAGAGCT
GAGTCTGAGCCACGTGGCAGACCAAATGATCGGCAACTATAATTTTGGGGGGAT
TTCCAGTGGCGAGCGGCGCCGAGTGTCCATCGCAGCCCAACTCCTTCAGGACCC
CAAGGTCATGATGCTTGACGAGCCAACCACAGGACTGGACTGCATGACTGCAA
ATCATATCGTCCTCCTCTTGGTCGAGCTGGCTCGCAGGAACCGCATTGTAATTGT
CACCATCCACCAGCCTCGCTCTGAGCTCTTCCACCACTTCGACAAAATTGCCATT
CTGACTTACGGAGAGTTGGTGTTCTGTGGCACGCCAGAGGAGATGCTCGGCTTC

TTCAATAACTGTGGTTACCCCTGTCCTGAACATTCCAATCCCTTTGATTTCTACA
TGGACTTGACATCGGTGGACACCCAAAGCAGAGAGCGAGAGATAGAGACGTAC
AAGCGAGTCCAGATGCTGGAATCTGCCTTCAGGCAATCGGACATCTGTCACAAA
ATCCTGGAGAACATTGAAAGAACAAGACACCTGAAAACCCTACCCATGGTTCCT
TTCAAAACGAAAAATCCTCCCGGAATGTTCTGCAAGCTCGGCGTTCTCCTGAGG
AGAGTAACGAGAAACCTAATGAGGAATAAGCAGGTGGTGATTATGCGTCTTGTT
CAGAATCTGATCATGGGTCTGTTCCTCATTTTCTACCTTCTCCGAGTCCAGAACA
ACATGCTGAAGGGCGCTGTTCAGGACCGCGTAGGGCTGTTGTACCAGCTTGTGG
GTGCCACCCCGTACACCGGCATGCTCAACGCTGTGAACCTCTTTCCCATGCTGA
GAGCTGTCAGCGACCAGGAGAGTCAGGATGGCCTGTACCAGAAGTGGCAGATG
CTGCTCGCCTATGTGCTGCATGCTCTCCCCTTCAGCATCGTTGCCACGGTGATTT
TCAGCAGCGTGTGTTACTGGACTCTGGGCTTGTATCCCGAGGTCGCCAGATTTG
GATACTTCTCTGCCGCTCTGTTGGCCCCTCACTTAATTGGAGAATTTCTGACACT
TGTGCTGCTTGGTATGGTCCAAAACCCCAATATTGTCAACAGCATAGTGGCTCT
GCTGAGTATTTCTGGGTTGCTCATTGGATCTGGATTTATCAGAAACATAGAAGA
AATGCCCATTCCTTTAAAAATCCTGGGTTACTTTACCTTCCAAAAGTACTGTTGT
GAGATTCTTGTGGTCAATGAGTTCTATGGCCTGAACTTCACTTGTGGTGGCTCCA
ACACTTCTGTGCCAAATAACCCAATGTGTTCCATGACCCAAGGGATCCAATTCA
TTGAGAAAACCTGCCCAGGGGCCACGTCCAGATTCACGACAAACTTCCTGATCT
TGTACTCGTTCATCCCGACTCTTGTCATCCTGGGGATGGTGGTCTTTAAAGTCCG
GGACTACCTGATTAGCAGATAGGTAAGATGGCAGGCAGGAAAGGGTTAATGGG
CAGGCTCGCCCACTGTGGAGCACAGAGAAGTACAAGCC

SEQ. ID NO. 49

Hamster ABCG5 partial amino acid sequence

AISGRLRRTGTLEGEVFVNGRELRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAML
ALRSSSSDFYDKKVEAVMEELSLSHVADRMIGNYNFGGISSGERRRVSIAAQLIQDP
KIMMFDEPTTGLDCMTANQIVILLAELARRDRIVIVTIHQPRSELFQHFDKIAILTYGE
MVFCGTPEEMLDFFNSCGYPCPEHSNPFDFYMDLTSVDTQSREREIETYKRVQMLE
SAFRDSAVCHKILENIERTKHLKTLPMIPFKTKDPPGMFCKLGVLLRRVTRNLMRNK
QAVIMRLVQNLIMGLFLIFYLLRVQNDILKGAIQDRVGLLYSWSAPPRTPACST

SEQ. ID NO. 50

Hamster ABCG5 partial cDNA sequence

TCAGGCTCAGGGAAAACCACGTTGCTGGTGCCATCTCCGGGAGGCTGCGACGCA
CAGGGACCCTGGAAGGGGAGGTGTTTGTGAACGGCCGTGAGCTGCGCAGGGAC
CAGTTCCAAGACTGCTTCTCCTATGTCCTGCAGAGCGACGTCTTTCTGAGCAGTC
TCACGGTGCGAGAGACGCTGCGCTACACGGCGATGCTGGCCCTCCGCAGTAGCT
CTTCGGACTTCTATGACAAGAAGGTAGAGGCAGTCATGGAAGAGCTAAGTCTG
AGCCACGTGGCAGACCGAATGATTGGCAACTATAATTTTGGGGGAATTTCCAGT
GGCGAGCGGCGCCGAGTCTCCATCGCAGCCCAACTCATTCAGGACCCCAAGATC
ATGATGTTTGATGAGCCAACCACAGGACTGGACTGCATGACTGCAAATCAAATT
GTCATCCTCCTGGCAGAGCTGGCTCGCAGGGACCGCATTGTGATCGTCACCATC
CACCAGCCTCGCTCTGAGCTCTTTCAACACTTCGACAAAATTGCCATCCTGACTT
ACGGAGAGATGGTGTTCTGTGGCACGCCGGAGGAAATGCTCGACTTCTTCAATA
GCTGTGGTTACCCTTGTCCTGAACATTCCAACCCCTTTGACTTCTACATGGACTT
GACATCAGTGGATACCCAGAGCAGAGAGCGAGAAATAGAAACCTACAAGAGA
GTCCAGATGCTCGAATCTGCCTTCAGAGACTCTGCAGTCTGTCACAAAATCCTG
GAGAATATTGAAAGGACAAAACACCTGAAAACCTTACCCATGATTCCTTTCAAA
ACGAAAGATCCTCCTGGAATGTTCTGTAAGCTGGGTGTCCTCTTGAGGAGAGTT
ACAAGAAACTTAATGAGAAACAAGCAGGCAGTGATCATGCGTCTTGTTCAGAA
TCTCATCATGGGTCTGTTCCTCATTTTCTACCTTCTTCGGGTCCAGAACGACATA
CTAAAGGGCGCTATCCAGGACCGTGTGGGTCTGCTATACAGCTGGTCGGCGCCA
CCCCGTACACCGGCATGCTCAACGCTGTGAATTTGTTTCCCATG

Incorporation by Reference

Throughout this application, various publications, patents, and/or patent applications are referenced in order to more fully describe the state of the art to which this invention pertains. The disclosures of these publications, patents, and/or patent applications are herein incorporated by reference in their entireties to the same extent as if each independent publication, patent, and/or patent application was specifically and individually indicated to be incorporated by reference.

Other Embodiments

It will be apparent to those of ordinary skill in the art that various modifications and variations can be made in the present invention without departing from the scope or spirit of

the invention. Other embodiments of the invention will be apparent to those of ordinary skill in the art from consideration of the specification and practice of the invention disclosed herein. It is intended that the specification and examples be considered as exemplary only, with a true scope and spirit of the invention being indicated by the following claims.

5

WHAT IS CLAIMED IS:

1. A method of identifying a subject having a predisposition for developing sitosterolemia, comprising detecting a mutant ABCG5 polypeptide or a mutated ABCG5 nucleic acid in the subject, thereby identifying a subject having a predisposition for developing sitosterolemia.

2. A method of identifying a subject having a predisposition for developing arteriosclerosis or heart disease, comprising detecting a mutant ABCG5 polypeptide or a mutated ABCG5 nucleic acid in the subject, thereby identifying a subject having a predisposition for developing arteriosclerosis or heart disease.

3. The method of claim 1 or 2, wherein the mutated ABCG5 nucleic acid comprises a missense mutation.

4. The method of claim 1 or 2, wherein the mutated ABCG5 nucleic acid comprises a nonsense mutation.

5. The method of claim 1 or 2, wherein the mutated ABCG5 nucleic acid comprises a deletion mutation.

6. The method of claim 3, wherein a mutant ABCG5 polypeptide encoded by said mutated ABCG5 nucleic acid comprises a mutation at amino acid position 389.

7. The method of claim 6, wherein a mutant ABCG5 polypeptide encoded by said mutated ABCG5 nucleic acid comprises a histidine residue at amino acid position 389.

8. The method of claim 3, wherein a mutant ABCG5 polypeptide encoded by said mutated ABCG5 nucleic acid comprises a mutation at amino acid position 419.

9. The method of claim 8, wherein a mutant ABCG5 polypeptide encoded by said mutated ABCG5 nucleic acid comprises a histidine residue at amino acid position 419.

10. The method of claim 8, wherein a mutant ABCG5 polypeptide encoded by said mutated ABCG5 nucleic acid comprises a proline residue at amino acid position 419.

11. The method of claim 3, wherein a mutant ABCG5 polypeptide encoded by said mutated ABCG5 nucleic acid comprises a mutation at amino acid position 146.

12. The method of claim 11, wherein a mutant ABCG5 polypeptide encoded by said mutated ABCG5 nucleic acid comprises a glutamine at amino acid position 146.

13. The method of claim 4, wherein a mutant ABCG5 polypeptide encoded by said mutated ABCG5 nucleic acid terminates at amino acid position 243.

14. The method of claim 4, wherein a mutant ABCG5 polypeptide encoded by said mutated ABCG5 nucleic acid terminates at amino acid position 408.

15. The method of claim 5, wherein a mutant ABCG5 polypeptide encoded by said mutated ABCG5 nucleic acid is deleted of exon 3.

16. A method of identifying a mutant ABCG5 polypeptide or a mutated ABCG5 nucleic acid encoding said mutant ABCG5 polypeptide, said polypeptide having reduced selectivity for internalization of non-cholesterol sterol in an intestine or hepatic cell, comprising detecting, in a patient with sitosterolemia, a ABCG5 polypeptide that is not present in normal subjects or an ABCG5 nucleic acid that is not present in normal subjects, thereby identifying a mutant ABCG5 polypeptide or a mutated ABCG5 nucleic acid encoding said polypeptide having reduced selectivity for internalization of non-cholesterol sterol in an intestine or hepatic cell.

17. A method of identifying a compound which alters ABCG5 activity level, comprising:

contacting a cell culture comprising an ABCG5 polypeptide with a compound; and

measuring ABCG5 biological activity in the cell culture,

whereby an increase in ABCG5 biological activity compared to ABCG5 biological activity in a control cell culture not contacted with the compound, identifies a compound which increases ABCG5 biological activity, or,

whereby a decrease in ABCG5 biological activity compared to ABCG5 biological activity in a control cell culture not contacted with the compound, identifies a compound which decreases ABCG5 activity.

18. The method of claim 17, wherein said cell culture comprises cells comprising a mutant ABCG5 polypeptide.

19. A method of identifying a compound which alters ABCG5 biological activity level, comprising:

contacting a mammal having cells comprising an ABCG5 polypeptide with a compound; and

measuring ABCG5 biological activity in the mammal,

whereby an increase in ABCG5 biological activity compared to ABCG5 biological activity before contacting the mammal with the compound, identifies a compound which increases ABCG5 activity, or,

whereby a decrease in ABCG5 biological activity compared to ABCG5 biological activity before contacting the mammal with the compound, identifies a compound which decreases ABCG5 activity.

20. The method of claim 17 or 19, wherein said cell in said cell culture or mammal comprises a mutated ABCG5 polypeptide or a wild type polypeptide.

21. The method of claim 20, further comprising comparing said ABCG5 biological activity, or level of ABCG5 mRNA, or level of ABCG5 polypeptide in the cell culture or mammal to ABCG5 biological activity, or level of ABCG5 mRNA, or level of ABCG5 polypeptide in a second cell culture or mammal comprising a wild type ABCG5 polypeptide.

22. A method of modulating transport of a sterol by a cell, comprising modulating ABCG5 biological activity in the cell, thereby modulating transport of the sterol by the cell.

23. The method of claim 22, wherein the sterol is phytosterol.

24. The method of claim 22, wherein the sterol is cholesterol.

25. The method of claim 22, wherein the sterol is sitosterol.

26. The method of claim 22, wherein ABCG5 biological activity is increased.

27. The method of claim 26, wherein ABCG5 biological activity is increased by increasing the amount of functional ABCG5 polypeptide within the cell.

28. The method of claim 26, wherein transport of the sterol is increased.

29. The method of claim 26, wherein excretion of the sterol from the cell is increased.

30. A method of increasing sterol excretion in a subject, comprising increasing ABCG5 biological activity in a hepatocyte in the subject, thereby increasing sterol excretion in the subject.

31. A method of decreasing sterol absorption in a subject, comprising increasing ABCG5 biological activity in an intestinal cell in the subject, thereby decreasing sterol absorption in the subject.

32. A method for improving the prognosis or ameliorating a disease state selected from the group consisting essentially of breast cancer, coronary heart disease, acute thrombosis, and stroke, comprising

administering to a patient an agent which decreases ABCG5 biological activity and results in increased sitosterol levels in said patient.

33. The method of claim 32, wherein the increase in sitosterol levels is to at least about 20% relative to sitosterol levels expected or observed for that patient prior to administration of said agent.

34. The method of claim 32, wherein the increase is between about 30% and 50%.

35. An isolated nucleic acid encoding ABCG5.

36. The isolated nucleic acid of claim 35, wherein the nucleic acid encodes mammalian ABCG5.

37. The isolated nucleic acid of claim 36, wherein the mammalian ABCG5 is human ABCG5.

38. The isolated nucleic acid of claim 37, wherein the nucleic acid comprises the nucleotide sequence set forth in SEQ ID NO: 41.

39. The isolated nucleic acid of claim 35, wherein the nucleic acid comprises a nucleotide sequence that encodes a mutant ABCG5 polypeptide.

40. The isolated nucleic acid of claim 39, wherein the nucleotide sequence encodes a mutant ABCG5 polypeptide comprising a mutation at amino acid position 145, 243, 389, 408, 419, or is missing exon 3.

41. A vector comprising a nucleic acid encoding ABCG5.

42. The vector of claim 41, wherein the nucleic acid encodes mammalian ABCG5.

43. The vector of claim 41, wherein the ABCG5 nucleic acid is operably linked to a transcriptional promoter.

44. A non-human transgenic mammal comprising an isolated nucleic acid encoding mammalian ABCG5.

45. The non-human transgenic mammal of claim 44, wherein the non-human transgenic mammal is a mouse.

46. The non-human transgenic mammal of claim 44, wherein the nucleic acid encodes human ABCG5.

47. A non-human mammal comprising a deleted, mutated, or polymorphic variant heterozygous ABCG5 gene.

48. The non-human mammal of claim 47, wherein the non-human mammal is a mouse.

49. The non-human mammal of claim 47, wherein the non-human mammal encodes a human ABCG5 gene.

50. An isolated mammalian ABCG5 polypeptide.

51. The isolated polypeptide of claim 50, wherein the mammalian ABCG5 polypeptide is a human ABCG5 polypeptide.

52. The isolated ABCG5 polypeptide of claim 51, wherein the polypeptide comprises the amino acid sequence set forth in SEQ ID NO: 40.

53. The isolated ABCG5 polypeptide of claim 51, wherein the polypeptide comprises an amino acid sequence that is a mutant ABCG5 polypeptide.

54. An isolated antibody that specifically binds an ABCG5 polypeptide.

55. The isolated antibody of claim 54, wherein the ABCG5 polypeptide is a human, mutated ABCG5 polypeptide.

56. The isolated antibody of claim 55, wherein the isolated antibody is a polyclonal antibody.

57. The isolated antibody of claim 55, wherein the isolated antibody is a monoclonal antibody.

58. An isolated dimer half-transporter enzyme comprising at least one ABCG5 monomer.

59. The isolated dimer half-transporter enzyme of claim 58, wherein at least said ABCG5 monomer is a human ABCG5 polypeptide.

FIG. 1

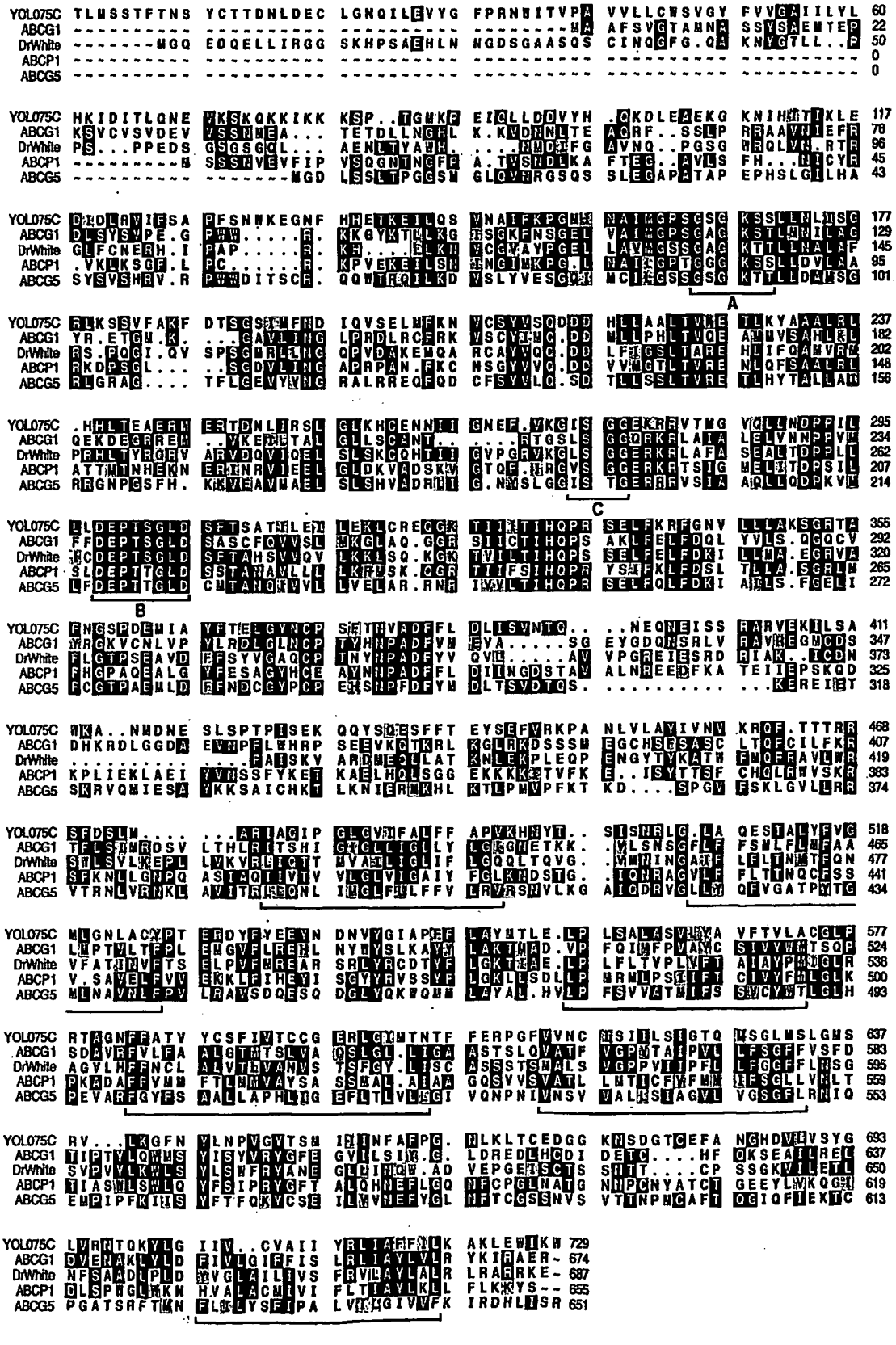


FIG. 2

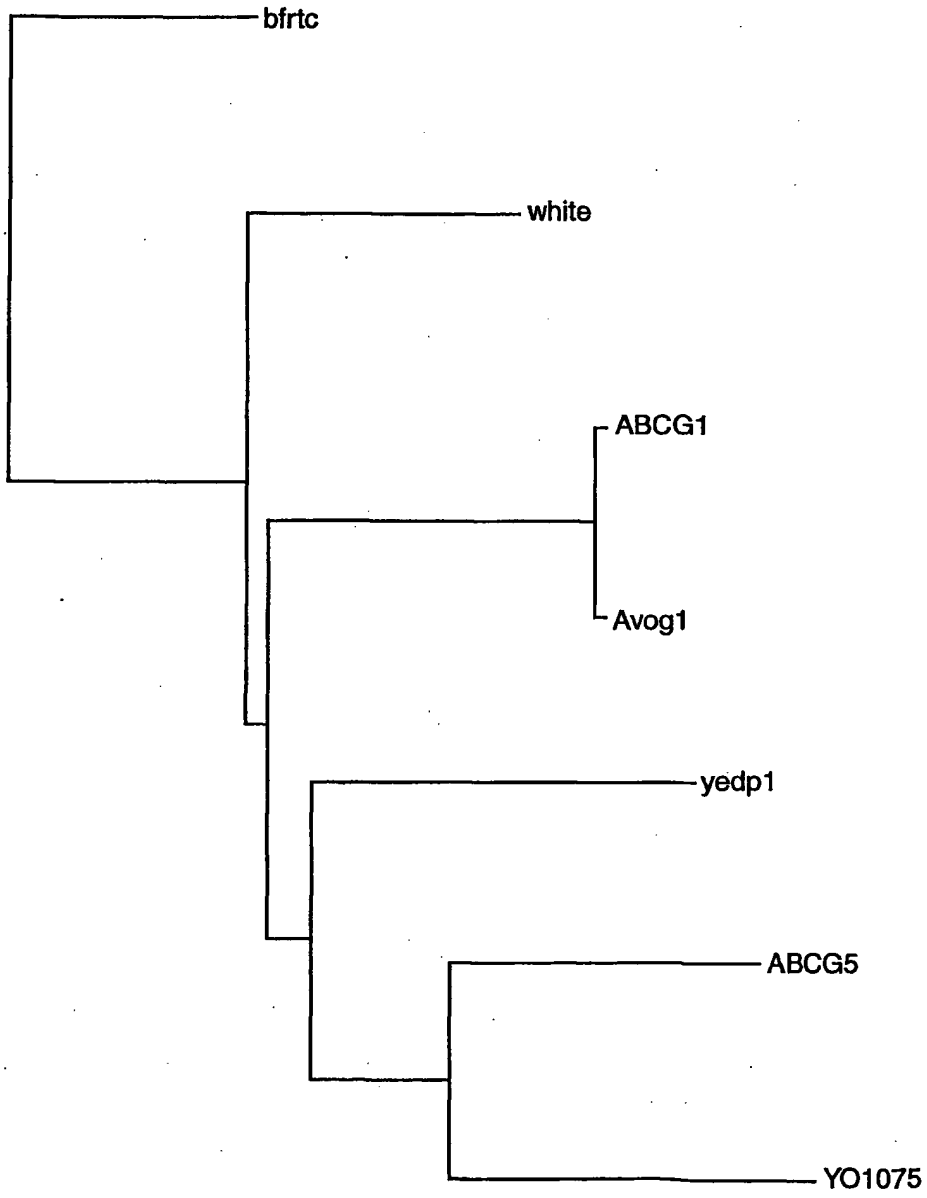


FIG. 3

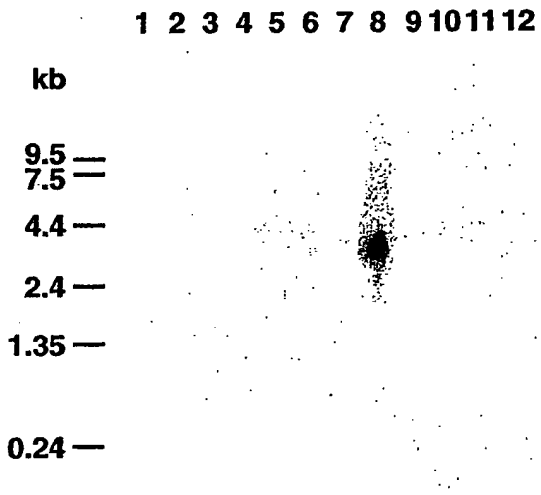


FIG. 4

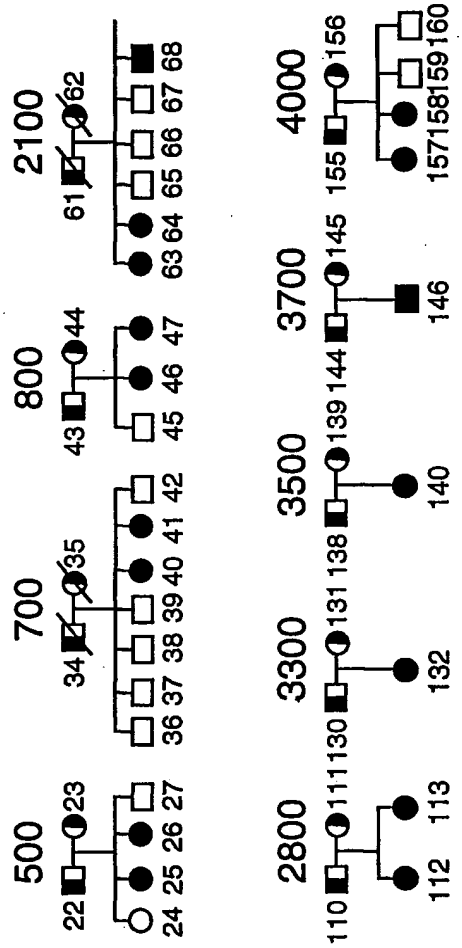


FIG. 5A

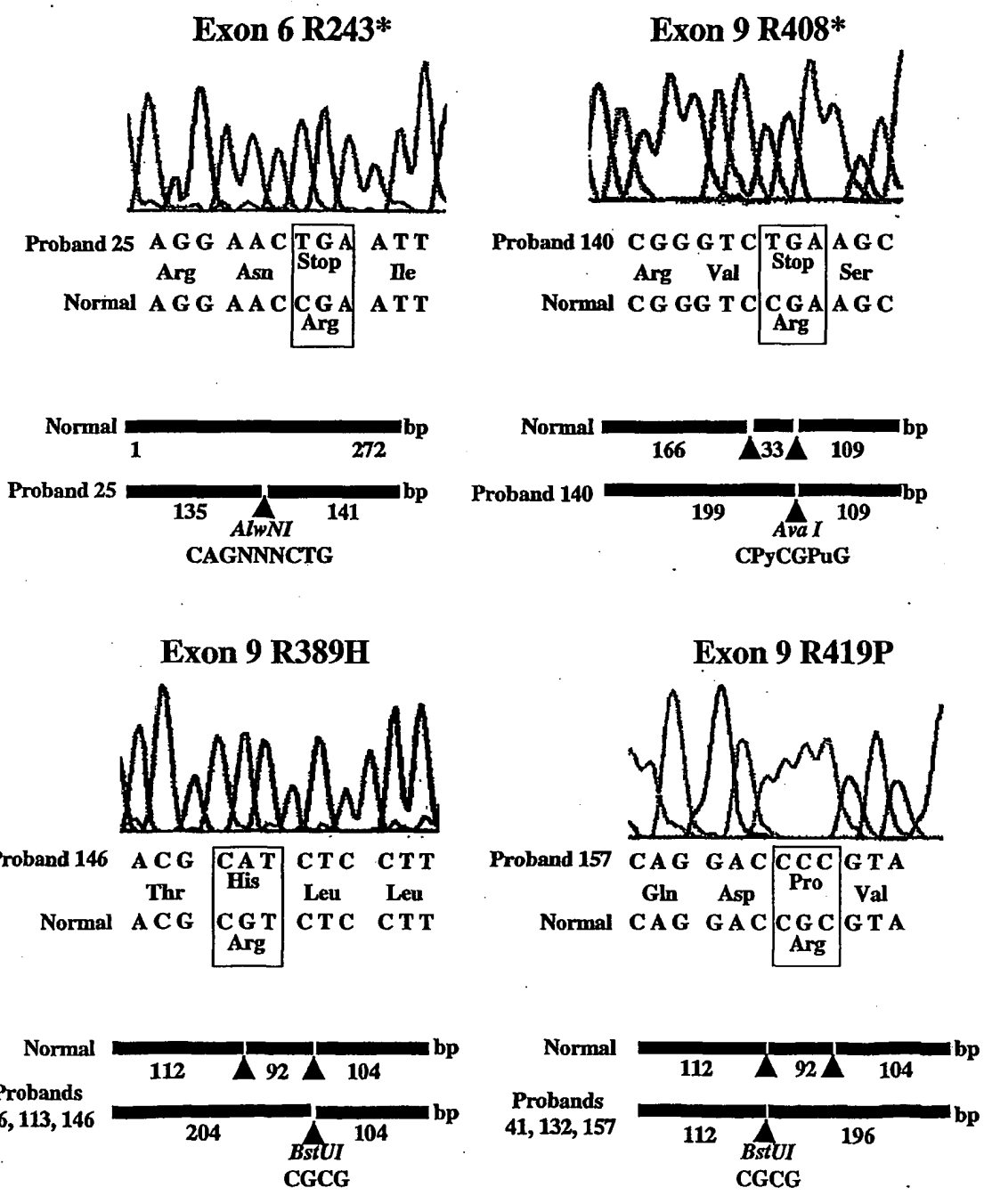

Exon 6 R243*
Proband 25 A G G A A C T G A A T T
Arg Asn Stop Ile
Normal A G G A A C C G A A T T
Arg
Normal bp
1 272
Proband 25 bp
135 141
AlwNI
CAGNNNCTG
Exon 9 R408*
Proband 140 C G G G T C T G A A G C
Arg Val Stop Ser
Normal C G G G T C C G A A G C
Arg
Normal bp
166 33 109
Proband 140 bp
199 109
Ava I
CPyCGPuG
Exon 9 R389H
Proband 146 A C G C A T C T C C T T
Thr His Leu Leu
Normal A C G C G T C T C C T T
Arg
Normal bp
112 92 104
Probands 46, 113, 146 bp
204 104
BstUI
CGCG
Exon 9 R419P
Proband 157 C A G G A C C C C G T A
Gln Asp Pro Val
Normal C A G G A C C G C G T A
Arg
Normal bp
112 92 104
Probands 41, 132, 157 bp
112 196
BstUI
CGCG

FIG. 5B

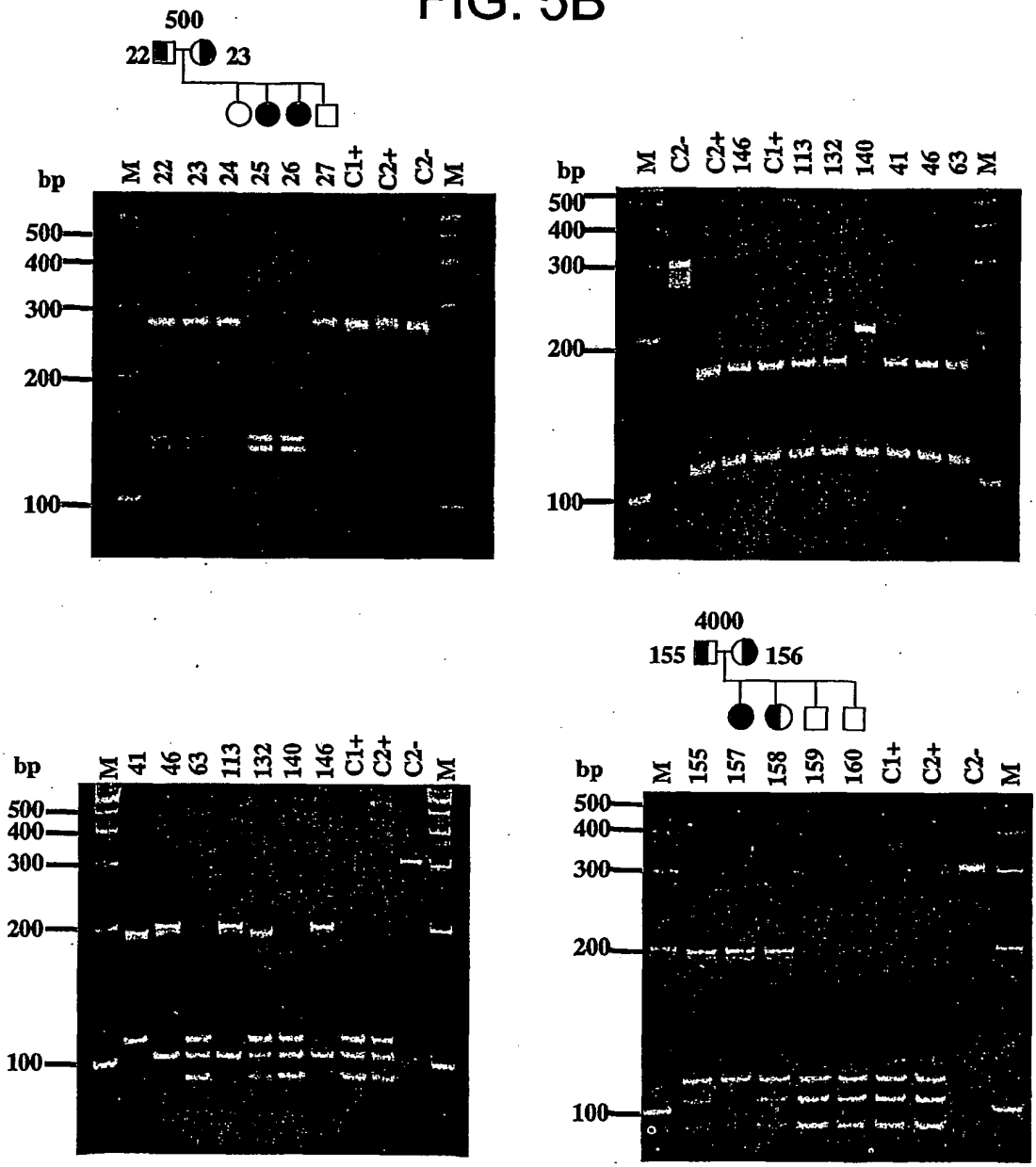

500
22
23
bp
M 22 23 24 25 26 27 C1+ C2+ C2- M
500
400
300
200
100
bp
M C2- C2+ 146 C1+ 113 132 140 41 46 63 M
500
400
300
200
100
4000
155
156
bp
M 41 46 63 113 132 140 146 C1+ C2+ C2- M
500
400
300
200
100
bp
M 155 157 158 159 160 C1+ C2+ C2- M
500
400
300
200
100

FIG. 6

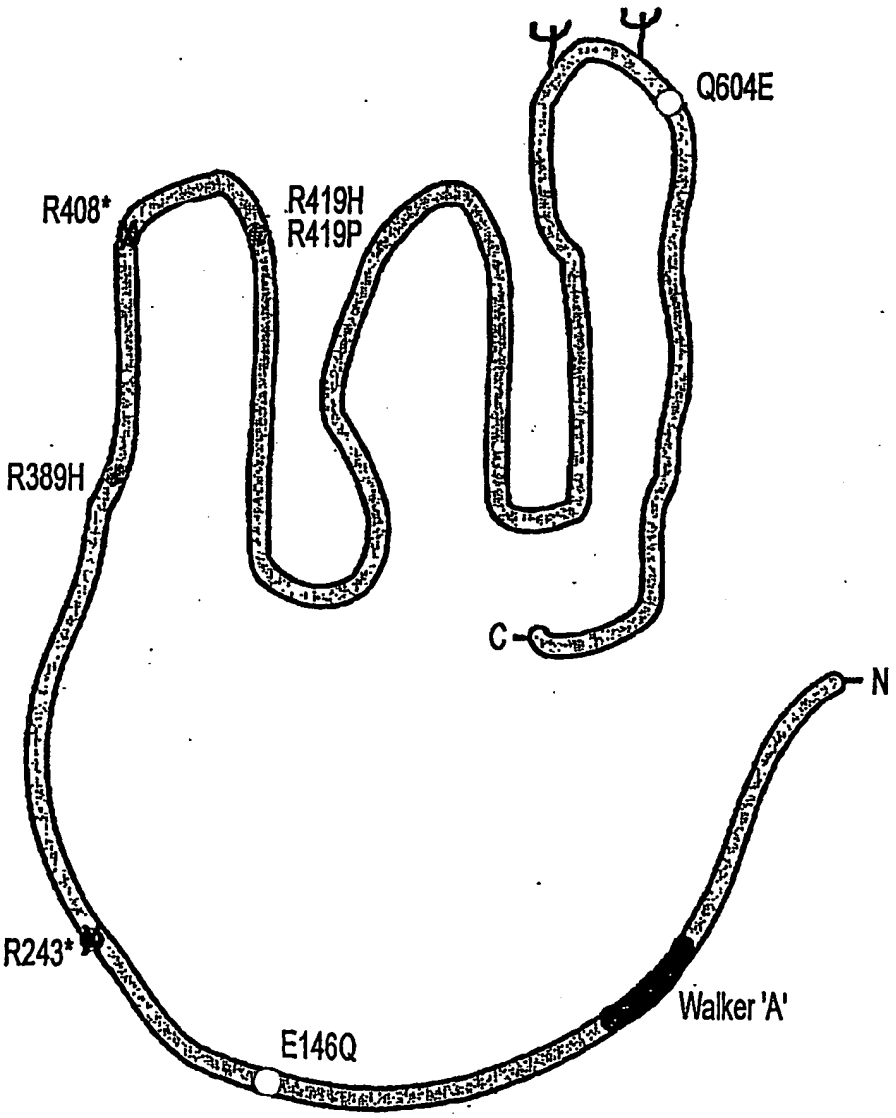


FIG. 7

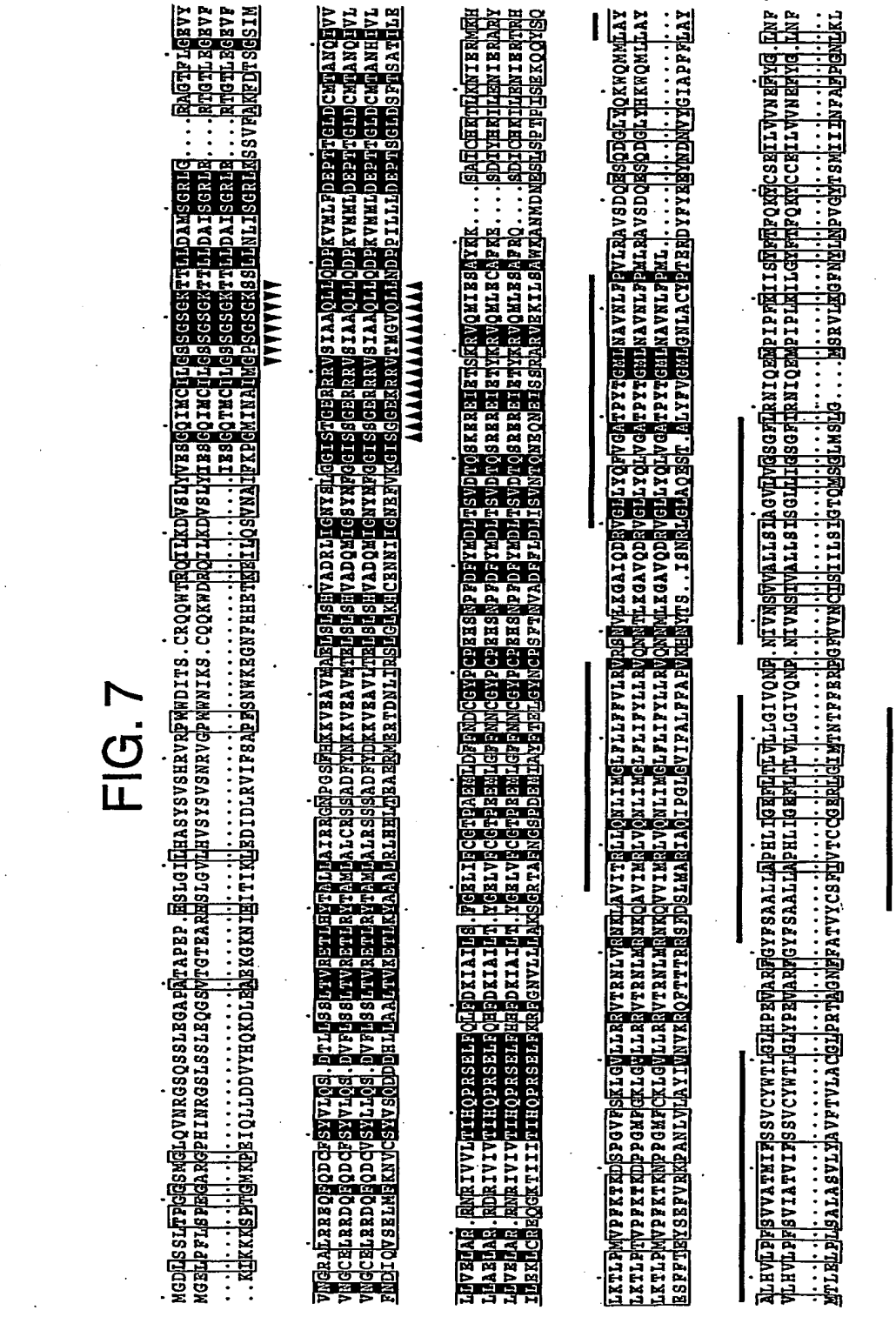


FIG. 8

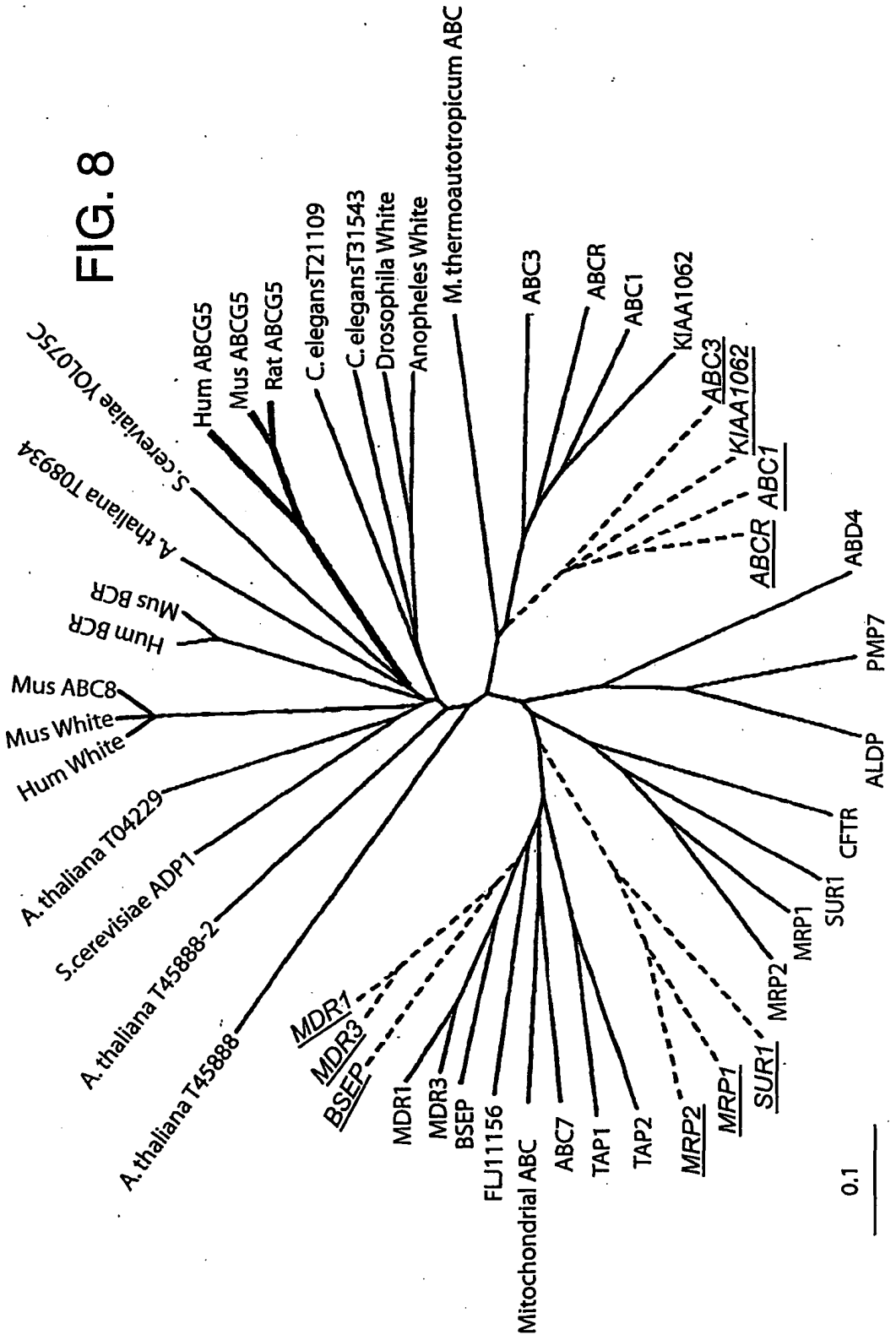
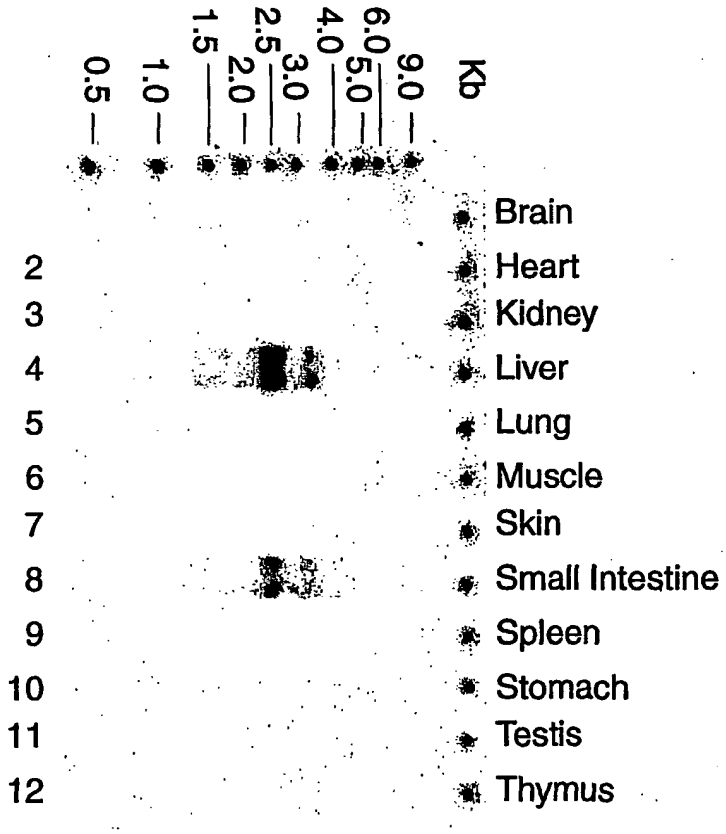

M.thermoautotropicum ABC
Anopheles White
Drosophila White
C.elegansT31543
C.elegansT21109
Rat ABCG5
Mus ABCG5
Hum ABCG5
S.cerevisiae YOL075C
A.thaliana T08934
Mus BCR
Hum BCR
Mus ABC8
Mus White
Hum White
A.thaliana T04229
S.cerevisiae ADP1
A.thaliana T45888-2
A.thaliana T45888
MDR1
MDR3
BSEP
FLJ11156
Mitochondrial ABC
ABC7
TAP1
TAP2
MRP2
MRP1
SUR1
MRP2
MRP1
SUR1
CFTR
ALDP
PMP7
ABD4
ABCR
ABC1
KIAA1062
ABC3
ABC3
ABCR
ABC1
KIAA1062
0.1

FIG. 9


0.5
1.0
1.5
2.0
2.5
3.0
4.0
5.0
6.0
9.0
Kb
Brain
Heart
Kidney
Liver
Lung
Muscle
Skin
Small Intestine
Spleen
Stomach
Testis
Thymus
2
3
4
5
6
7
8
9
10
11
12